MATERNAL LINEAGE EFFECTS AND GENETIC DIVERSITY IN THE UK DAIRY POPULATION

Tim Roughsedge

BSc. (Hon.) University of Newcastle upon Tyne 1993

Ph.D.

University of Edinburgh

1999

CONTENTS

DECLARATION

I declare that this thesis has been composed by me.

Specific contributions of others are acknowledged.

T. Roughsedge

November 1999

ACKNOWLEDGEMENTS

I thank Peter Visscher and Sue Brotherstone for supervising my thesis work. I have always been able to come to either of you for discussions, often at very short notice if any was given and you have both made time for me. I have always felt very safe in the knowledge that you were both there to turn to.

I thank the Ministry of Agriculture Food and Fisheries in the UK for their financial support that funded this study. I also thank the organisers of EAAP, BSAS and MAFF for at various times supporting my visits to some very interesting scientific meetings. A special thanks to HUKI for use of their datbase and NMR for allowing me to use their data.

Thanks to Bill Hill for providing valuable comments on papers based on Chapters 2 and 3. I also thank Roel Veerkamp for being my third supervisor (for a very short time) and for making comments on a paper based on Chapter 2.

I would also like to thank my office mates past and present for putting up with me, including Rob Berry (the Led Zep years) through to Eileen Wall (the STEPS years).

I also extend my thanks to my flatmates Craig and Owen and for being there to try and keep me calm and put the world to rights while watching Emmerdale and the Big Breakfast.

Esther for keeping me going and occasionally reminding me that it is all going to be OK.

And last but by no means least my family for always supporting my every move even if they are not interested in my re-prints.

Peter Visscher made suggestions about the construction of the ANOVA table in Chapter 6 and also proposed the test statistic that was used in Chapters 2, 4 and 5.

PUBLICATIONS

Papers submitted from this thesis are:

Roughsedge T., Brotherstone S., Visscher P.M. (1999). Effects of cow families on type traits in dairy cattle. *Animal Science*. Submitted.

Roughsedge T., Brotherstone S., Visscher P.M. (1999). Effects of cow families on production traits in dairy cattle. *Animal Science*. Submitted.

Published papers (in appendix):

Roughsedge T., Brotherstone S., Visscher P.M. (1998). Lack of evidence for cytoplasmic inheritance for milk production traits at the Langhill dairy herd. *Proceedings of the 6th world congress on genetics applied to livestock production, Armidale*. **23**: 351-354.

Roughsedge T., Brotherstone S., Visscher P.M. (1999). Quantifying genetic contributions to a dairy cattle population using pedigree analysis. *Livestock Production Science*. **60**:359-369.

Roughsedge T., Brotherstone S., Visscher P.M. (1999). Estimation of variance of maternal lineage effects at the Langhill dairy herd. *Animal Science* **68**: 79-86.

Roughsedge T., Brotherstone S., Visscher P.M. (1999). Using pedigree analysis to determine the genetic diversity of the British dairy population over the last thirty years. BSAS winter meeting 1999.

ABSTRACT

There is a perceived importance of maternal families among dairy cattle breeders. The aim of this study was to estimate the magnitude of the component of phenotypic variance attributable to maternal lineage for conformation and production traits in the UK Holstein Friesian dairy cow population. Various population parameters were also estimated.

Advanced statistical methodology including restricted maximum likelihood was used to estimate maternal lineage variance. Production traits at the Langhill dairy herd were analysed to provide insight into feed intake and efficiency traits. Only fat yield had a significant component of phenotypic variance attributable to maternal lineage. In this trait 4% of the variance was attributable to maternal lineage. Production records of over 50,000 cows from the UK Holstein Friesian population were used to estimate maternal lineage variance with a contemporary data structure. This data structure was used to reduce the average within maternal family nuclear genetic relationship while maintaining the same expected magnitude of maternal lineage variance.

Dairy cattle breeders stress the importance of cow families in breeding for good conformation. To investigate this, type classification records of 32,000 heifers were analysed and maternal lineage variances estimated for all 23 conformation traits. Significance of the maternal lineage component was determined using a likelihood ratio test. Most researchers use an incorrect test when determining the significance level of a component. An explanation of the correct test is given. Principal component analysis was used to determine the number of independent components accounting for the variance in type traits. This number was then used to establish the significance level of the variance component test statistic. The only type traits to show significant effects were stature, a

linear type trait, and body, a composite trait. These traits had maternal lineage variance components of 1.0% and 1.5 % respectively.

Population parameters of the UK Holstein Friesian cow population were estimated. Conservation biology parameters demonstrated that, when 1960 was treated as a pseudo founder generation, only an equivalent of 1% of the founders were responsible for the genomic diversity of the 1997 population. It was shown that the introduction of large numbers of Holsteins from North America over recent years has reduced the level of inbreeding but at the same time reduced the genetic diversity of the population. It was found that in 1997 one ancestor accounted for 5% of the genome of the UK Holstein Friesian cow population. Average degree of relationship was also shown to be increasing at a rate of 0.07% per year. It can be hypothesised that in the future the global Holstein Friesian population will show an increase in average inbreeding coefficient and average pairwise degree of relationship.

A theoretical investigation was made of the consequences of incorrect maternal family assignment, through both pedigree errors and the inadequate tracing of pedigrees, on the magnitude of the variance component estimated. This demonstrated that the under-estimation of maternal lineage variance occurs unless complete family information is available and accurately recorded.

Chapter 1

A synthesis of cytoplasmic inheritance in dairy cattle: theoretical background and practical estimation of maternal lineage variance

1.1 Introduction

Over recent years interest has been expressed in the possible contribution of cytoplasmic inheritance to the phenotypic variance of economically important traits in dairy cattle. One of the first authors to draw attention to this possibility was Wagner (1972). Wagner (1972) reviewed the evidence for cytoplasmic inheritance in plant species and yeast but provided no evidence from animal species to support cytoplasmic inheritance. The question raised was, 'how would it be possible to test the hypothesis of cytoplasmic inheritance occurring in animals?'

The estimation of breeding values in dairy cattle is based on knowledge of the component parts that contribute to the phenotypic variation of traits. At present the genetic component used in this evaluation is the additive genetic variance which is used to determine the heritability of a trait. However there has always been a belief amongst cattle breeders that certain cow families produce better cows than bulls in terms of genetic merit for production and conformation. There is also evidence to show that heritability estimates from daughter dam regressions are consistently higher than those obtained from paternal half-sib estimation (e.g. Seykora and McDaniel, 1983; Visscher and Thompson, 1992). This evidence may point towards a mechanism of inheritance, in addition to nuclear genetic inheritance, being transmitted through the female line, which is not being accounted for by current evaluations. One possible explanation for this is the inheritance of mitochondrial DNA (mtDNA). If maternal lineages are responsible for a component of variance that is not being accounted for in current breeding value estimations in dairy cattle then it is important that the effect is identified and quantified.

1.2 Basis for mitochondrial DNA being responsible for cytoplasmic inheritance

1.2.1 Physical structure

Mitochondria in cell cytoplasm carry their own DNA (deoxyribonucleic acid), which in the bovine is a 16338 nucleotide closed loop double helix. MtDNA contains 13 protein-coding genes, 12 of which are involved in the electron transport chain essential for converting ADP to ATP to provide energy for cellular function. It also contains 22 transfer ribonucleic acid (RNA) genes, and 2 ribosomal RNA genes (Anderson *et al.*, 1982). The mtDNA is composed of a light and heavy strand. The heavy strand codes for both ribosomal RNA genes and all but one of the proteins. A portion of the mtDNA called the displacement loop (D-loop) contains 910 base pairs. The promoters of the light and heavy strand are located in the D-loop, and the origin of heavy strand replication is in the D-loop. The D-loop is thought to be the least conserved region of mtDNA.

1.2.2 Basis for maternal lineage variance

In 1974 Hutchinson *et al.* (1974) demonstrated that mtDNA has an almost exclusively maternal route of inheritance in mammals. They took horses and donkeys and performed the reciprocal crosses, the outcome of which were hybrids that were sterile but demonstrated the route of mtDNA inheritance. The cross of the female horse with the male donkey resulted in a mule with detectable mtDNA of horse origin only. The same scenario was shown with the reciprocal cross, which resulted in a hinny with donkey mtDNA. The sensitivity of their test was not high and paternal mtDNA at a level below 5% would not have been detected. They suggested two possible mechanisms for mitochondrial DNA inheritance:

1) the paternal (spermatozoon) mitochondria may be incapable of replication during development of fertilised eggs.

2) there may just be a quantitative preponderance of egg derived mitochondria in the zygote.

A more sensitive test was performed by Gyllensten *et al.*, (1991). They took a backcrossing approach over 8 generations using distantly related mouse species to establish the inheritance of mtDNA. Evidence was provided to support the second mechanism proposed by Hutchinson *et al.* (1974) and mtDNA of paternal origin was found at a level of 10^{-4} relative to the maternal contribution in mice.

1.3 Reasons for investigating the contribution of mitochondrial DNA

The 37 genes encoded by mtDNA are few in comparison to the 50,000-100,000 that are encoded by nuclear plus mitochondrial DNA so why is it worth investigating? There are a number of reasons, and of primary importance is the nature of the process that the proteins coded by mtDNA are involved in. The mitochondria are the energy factories of cells and are therefore central to cell function. Gibson *et al.*, (1997) gave a theoretical approximation of the possible significance of mtDNA to phenotypic differences. Given a typical coefficient of variation (C.V.), for lactation traits of 18%, with mitochondria accounting for 1, 2, 5 and 10% of total phenotypic variance, the top 5% of lineages would exceed average performance by 3.7, 5.2, 8.3 and 11.7%.

1.4 Detection of a cytoplasmic lineage component

1.4.1 Pedigree considerations

In order to perform a statistical analysis to detect a component that can be attributed to the inheritance of mtDNA cows must first be assigned to maternal family groups. Pedigrees

of the cows need to be traced back to establish common ancestral links. The theory is that the mitochondria of each cow of a maternal family will be identical. Mutation rates in mitochondria are much greater than in nuclear DNA though this should not affect the small number of generations we are considering. The earliest traced points of convergence, a convergence being a common ancestor, in the analysis are taken as the points of cytoplasmic origin and define the lineages. There is a necessity to compromise several factors in this lineage assignment. Assigning more cows to a common lineage, which is achieved by tracing the pedigree further back in time and thus providing more convergence, enhances the power of detection of maternal lineage variance. However in tracing the pedigree further back there is also a greater chance that pedigree recording errors occur. Ron *et al.* (1996) summarised estimates of paternity misidentification rates of between 1.3 and 23 % across a number of European studies. Though this is a paternity estimate it would not be unreasonable to hypothesise that some degree of maternity misidentification occurs. Gibson *et al.* (1997) provide an example of the simplest case of this occurring. Assume that all recorded cows are an equal number of generations, *n,* from the founder female of the lineage, who is also the nearest common ancestor, ensuring that all errors of allocation of cows to the lineage are independent of each other. If we then take r as the recording error rate per generation, the proportion of females that are truly a part of the lineage is $(1 - r)^n$. The expected estimate of the lineage variance will be $[(1 - r)^n]^2$ of the true variance. For example, if we take a data structure of n = 6 generations deep, with an error rate of r = 0.01 per generation, then this gives us an estimate of lineage variance as a proportion of true variance of, $[(1-0.01)^6]^2 = 0.89$. If r were known then a correction factor could be fitted. The best method suggested for the determination of r is repeated simulation of the observed pedigree structure with random allocation of incorrect parentage at the predicted error rate. The effect that pedigree error rate has on the magnitude of the variance component estimated in relation to the true variance component is investigated in chapter 6.

1.4.2 Mitochondrial DNA variation

In order that cows can be assigned to cytoplasmic lineages based on maternal families it must be established that there is sufficient between family variation in mtDNA and low within family variation. Brown *et al.* (1988) used maternal families of the Iowa State University Dairy Breeding Herd to investigate variation in mtDNA at the molecular level The herd pedigree was traced to origin animals which resulted in 53 lineages, 38 of which had contemporaries to be used for molecular analysis. In their study only one cow per maternal lineage from 29 maternal lineages was used with the assumption of no within lineage variation. The D-loop region of mtDNA as already stated is thought to be the least conserved region and was therefore targeted in the molecular analysis. Using the restriction fragment length polymorphism technique Brown *et al.* (1988) established that 12 of the 29 maternal lineages targeted showed mtDNA sequence differences to the sequence published by Anderson *et al.* (1982). These differences were interpreted as single base changes and no evidence was found of gross insertions, deletions or re-arrangements. It was concluded that sufficient variation was shown to account for phenotypic variance. Freeman (1990) continued the work with the Iowa State University Dairy Breeding Herd. It was shown at that time that 51 separate sites of sequence variation had been identified in 38 lineages. Nucleotide substitutions were found at 48 sites; there were 9 base pair deletions, and 2 variable length poly G/poly C regions were found of 6-8 and 11-16 base pairs, respectively. Variation within family was observed at two sites: one G-to-C transversion at (nt) 363 and in the length of a poly G/poly C region that extends from (nt) 351 to 363. The variation at (nt) 363 occurred in 29% of the lineages and this site was considered hyper-variable. A cytoplasmic gene tree constructed from the sequence data revealed a bifurcation in the mtDNA of cattle into two main groups based on a transition at (nt) 169. This study provided further evidence that sufficient variation existed in mtDNA to confidently assign animals to a cytoplasmic lineage.

1.5 Estimation of Maternal Lineage Variance

1.5.1 The Sire Model

Over the last 14 years various analyses have been performed in order to detect cytoplasmic inheritance in the lactation traits of dairy cattle. Differing levels of success have been achieved and various suggestions have been put forward to explain the discrepancies.

Given a population with good pedigree records with a low assignment error rate so that maternal lineage can be traced, it is possible to fit effects or estimate variance components attributable to differences between maternal lineages. Having said that, one must be very careful in model choice and data structure when estimating maternal lineage variance. Salehi and James (1997), using a simulated data set, demonstrated how important it is to have a degree of pedigree depth. The data that they simulated related to a trait in a sheep population with varying numbers of years of information, and hence numbers of generations to cytoplasmic origin, and also variable size of data set. They found that the power of detection increased with the size of the cytoplasmic component simulated but was independent of the corresponding heritability component. By increasing the size of data set the power was also increased but given the same amount of data over a longer time period with more generations to the origin the power was further increased. In the case of dairy cattle populations there is not a big problem with pedigree depth though some of the analyses have been performed on experimental herds where the number of records available raises power issues. In one of the first attempts to estimate cytoplasmic effects in dairy cow production traits Bell *et al.* (1985) applied a sire model fitting maternal lineage as a fixed effect and performed the analysis using a least squares procedure. They looked at first lactation data on 4461 cows in five herds of the North

Carolina Department of Agriculture and one herd of North Carolina State University, calving between 1949 and 1980. The assignment of lineage involved a full trace of cows to the first female purchased outside of the herds being analysed, and a further trace if more than one founder female was purchased from the same breeder. After this assignment they then deleted all lineages from the analysis with less than 5 descendants. This process resulted in 102 cytoplasmic lines. With this approach they estimated cytoplasmic effects of 2.0, 1.8, 1.8, and 3.5% for milk yield, milk fat yield, 3.7% fat corrected yield and fat percentage. Based on their work they concluded that cytoplasmic lineage is a significant source of variation in lactation traits.

Kennedy (1986) investigated the model being used for analysis by simulating the parameters of the population used by Bell *et al.* (1985). Kennedy (1986) suggested that drift variance between lineages would be expected to accumulate in proportion to the number of generations from cytoplasmic origin. It is therefore intuitive to expect this additive drift variance to account for between lineage variance. Kennedy (1986) showed that using the same models as Bell *et al.* (1985) it was possible to estimate maternal lineage variance components of similar magnitude to the experimental study when none were in fact simulated. With the simulated data Kennedy (1986) was also able to fit the true additive genetic variance of the dam in the model and when he did so a zero maternal lineage variance component was estimated. This sort of analysis is not possible with experimental data where we do not know the true additive genetic value of the dam but it demonstrates the inadequacy of the sire model.

Huizinga *et al.* (1986) who looked at 290 first lactation records, which were assigned to 66 lineages, performed another early analysis. The cows, originally Dutch Friesian, were crossed with Dutch Holstein, Dutch Friesian or British Friesian bulls, in the Wageningen experimental herd. They performed a very similar analysis to that of Bell *et al.* (1985), using least squares methods and fitting lineage as a fixed effect. They found lineage to be

significant for milk yield, fat plus protein yield and milk returns, with implied contributions to total variance of 5.6, 10.1 and 12.5% respectively. However, the data was only three generations from the cytoplasmic origin. Therefore there were high nuclear additive genetic covariances within maternal lineage. The authors state that whilst this covariance was confounded with maternal lineage effects, the effect that they estimated was higher than could be explained by within lineage genetic covariance.

1.5.2 Theoretical Developments

Beavis *et al.* (1987) demonstrated a theoretical framework for separating the various components of phenotypic variance and their interactions. They were able to provide an experimental design to estimate cytoplasmic variance but this component was confounded with the additive by cytoplasmic interaction component. Laknath and Pollak (personal comm.) have recently demonstrated a way of overcoming this confounding and separating cytoplasmic variance using information on maternal half sibs and offspring maternal parent regression. This utilises the North Carolina Design II of Comstock and Robinson (1948) but the application to field data is not feasible. The model proposed by Beavis *et al.* (1987) was criticised by Kirkpatrick and Dentine (1988) for making rigid assumptions about the mode of inheritance. Kirkpatrick and Dentine (1988) proposed a path analysis approach looking at all possible variance components and the proportionate contribution made by each to the animal's phenotype and its offspring's phenotype. Reed and Van Vleck (1987) proposed an alternative method to estimate cytoplasmic variance given the criticism that Kennedy (1986) made of the method of Bell *et al.* (1985). In order to deal with the potential drift variance between lineages their approach was to use regression to estimate heritability with a daughter-dam design versus a granddaughter-granddam design. They suggested that this method would be able to identify cytoplasmic variance by obtaining higher heritability estimates from the granddaughter granddam estimate. They found no significant difference between the two estimates for milk yield traits to indicate a cytoplasmic component and concluded that cytoplasmic inheritance was not

responsible for a component of yield traits. This analysis can be criticised for fitting a herd year season effect for daughters only, and assigning all dams and granddams to a single herd year season, and thus making no correction to dams and granddams milk yield to account for their environment. The Kirkpatrick and Dentine (1988) model demonstrated that it was possible to have discrepancies in the estimates of heritability from daughter-dam and paternal half-sib analysis. They also showed that similar heritability estimates could be obtained from granddaughter granddam and daughter dam designs in the presence of cytoplasmic inheritance due to the presence of positive maternal genetic effects. The Kirkpatrick and Dentine (1988) model provides an alternative explanation for the results of Reed and Van Vleck (1987). In general there is a problem with 'competing' models, which all utilise alternative hypotheses to explain the data, e.g. cytoplasmic, sex-linked, maternal and imprinting effects. These effects would all produce unique patterns of inheritance in designed experiments but in the sort of data structure that is being used in dairy cattle analysis their effects would be largely indistinguishable.

1.5.3 The Animal Model

Kennedy (1986) concluded that the use of an animal model would properly separate the various components of phenotypic variance. The full implementation of an animal model to simultaneously estimate additive, maternal and cytoplasmic genetic effects was undertaken by Southwood *et al.* (1989). Data was simulated according to the procedure of Kennedy (1986) and in addition maternal genetic and cytoplasmic variance components were simulated. The data was then analysed with the animal model, correctly and incorrectly accounting for the components simulated. It was established that if the correct model was used then it was possible to estimate the true values of the components but if an incorrect model were applied then spurious variance components would be estimated. In all cases of incorrect model assignment the additive genetic variance component was inflated. If a maternal genetic variance component was simulated and then the data was analysed accounting only for a cytoplasmic component then a proportion of the maternal

genetic component was assigned to the cytoplasmic component but the majority of the extra variance inflated the additive component. Southwood *et al.* (1989) strongly encouraged the use of the animal model in order to estimate a cytoplasmic variance component. The animal model is used to estimate maternal lineage variance in Chapters 2, 4 and 5 of this study.

Bias in the estimation of cytoplasmic variance under the animal model was investigated recently (Rorato *et al.*, 1999). They re-analysed data used by Albuquerque *et al.* (1998) and also simulated data sets based on this data with randomly assigned levels to the cytoplasmic lines. Using restricted maximum likelihood (REML) to analyse the data under an animal model with cytoplasmic line as a random effect they obtained estimates of 0.3-0.4% of phenotypic variance attributable to cytoplasmic line for the simulated data. This upward bias of estimation was attributed to REML being forced to obtain positive variance estimates and is not unique to estimating a cytoplasmic effect. Rorato *et al.* (1998) also analysed the data using a sire model and confirmed that the sire model is not capable of separating additive and cytoplasmic line effects.

1.5.4 More recent analyses

In a continuation of the work outlined by Freeman (1990), Schutz *et al.* (1992) used the cows at the Iowa State University Dairy Breeding Herd to perform a phenotypic analysis. They used several approaches including a least squares procedure that treated lineage as a random effect. This approach detected significant lineage effects for milk yield, fat yield and fat%, explaining 4.1, 5.8 and 8.4% of total variance. They also performed an animal model best linear unbiased prediction (BLUP) procedure treating maternal lineage as a fixed effect and varying the heritability estimate used. In all cases they found significant lineage effects for fat %, but only for fat yield when a low value was used for heritability and no significant effect was found for milk yield. When the data were later reanalysed by

Schutz (as reported by Boettcher *et al.*, 1996a) using REML, and treating lineage as random in an animal model, no significant variance component was found to be attributable to cytoplasmic lineage for milk yield or the percentage of solids not fat traits, but 5.6 and 5.3% of the total phenotypic variance was explained by cytoplasmic lineage for fat% and energy concentration respectively.

Boettcher *et al.*, (1996a) analysed data of Iowa University State herd and 5 North Carolina herds using an animal model and fitting lineage as a random effect. They found results that agreed with previous analyses, in as far as little effect was found in connection with milk yield and protein yield traits and they found that fat% had the largest component attributable to cytoplasmic inheritance and was significant ($p=0.05$). However the effect for fat% was only 1.00% of the total phenotypic variance. The maternal lineage solutions to the equations for the traits considered were examined and the extreme solutions were found to be for the lineages with few individuals. When the data was reanalysed with size restricted to greater than 25 cows per lineage the solutions were much lower but little was altered in respect to the variance components.

In most analyses small data sets have been used for two main reasons. Firstly, the original analyses were performed on experimental herds (e.g. Bell *et al.*, 1985; Huizinga *et al.*, 1986; Boettcher *et al.*, 1996). Secondly, computational restraints did not allow the simultaneous estimation of several components of variance on large data sets. Boettcher and Gibson (1997) used a Bayesian approach to the estimation of cytoplasmic lineage variance in order to be able to deal with a very large data set. They used national production records of the Canadian Holstein population. After data editing they had 245,510 first lactation records distributed over 4 years. No effects greater than 0.5% were found for proportion of phenotypic variance attributable to maternal lineage. A small upward bias was detected in the estimation procedure though only of the magnitude of 0.1-0.2%. This upward bias was confirmed by the random assignment of maternal lineage

codes resulting in a 0.1-0.2% component. An analysis of field data was also undertaken by Albuquerque *et al.* (1998) using 68,063 cows from the population of Holsteins in New York. Data were spread over 11 years from 1980 to 1991. REML was used to estimate variance components from an animal model and for milk yield, fat yield and fat percentage the maternal lineage variance components estimated were 0.011, 0.008 and 0.009 respectively as a proportion of phenotypic variance.

All of the work so far discussed has been concerned with yield traits in dairy cattle. Schnitzenlehner and Essl (1999) looked at yield traits in the Austrian Simmental dairy cow population but in addition they also considered functional herd life and a measure of lactation persistency. This study used a large field data set analysing samples of ~10,000 records at a time. No consistent significant cytoplasmic variance component was found for any of the yield traits. Persistency and functional herd life however were estimated to have significant cytoplasmic variance components of 2.6 to 3.8% (depending on lactation number) for persistency and 4.6% for functional herd life. This gives us some indication that cytoplasmic inheritance plays a part in physiological mechanisms and may have a greater effect at certain times and hence the significance of persistency.

1 .5.5 Analysis Based On Sequence Information of Mitochondrial DNA

Work has mainly centered on the assignment of cows to a lineage with a common cytoplasmic origin in order to establish animals with common mitochondrial DNA. This approach is correct but relies on a number of assumptions. The main assumptions are that mtDNA is identical within lineage and all cytoplasmic origins are unrelated. The first of these assumptions will not hold true if there are assignment errors in pedigree recording. The result of this is downward biasing of the between lineage variance as previously discussed. Some recent work has allowed sequencing of mtDNA to be used to estimate relationships of specific genotypes. By using molecular markers to assign lineage the

problem of pedigree errors could be overcome and in addition the lack of knowledge in the development of maternal lineages could be removed. Schutz *et al.* (1993), using the same animals as Freeman (1990), fitted the sequence variations as a fixed effect in a phenotypic yield trait analysis using an animal model. This included fitting the main two groups defined as varying at nt 169, and also using clustering techniques based on the other sites of nucleotide variations to group similar lineages and fit these as effects. Some significant effects on fat percentage and milk energy were reported. The variation found in the hypervariable region and used by Schutz *et al.* (1993) is not variation in the coding region of mtDNA and can only provide a guide to there potentially being variation in the coding region. Schutz *et al.* (1994) found significant effects attributable to molecular variation in mtDNA with up to 842kg milk and 37kg fat differences due to single base-pair differences, though the authors add a note of caution in interpreting the results due to low numbers of cows in the study and the low frequency of some substitutions. Boettcher *et al.*, (1996b) used data from the breeding herd of Iowa State University and six North Carolina herds to examine the relationship between yield traits and mtDNA polymorphism. In the Iowa herd maternal ancestry was again used to assign cows to cytoplasmic lineages and it was assumed that mtDNA within a lineage was identical. Of the 81 lineages defined by this procedure only 29 were found to have surviving members to have their mtDNA sequenced. Eleven sites of polymorphism in the rRNA encoding sequences were identified. Polymorphism was detected by comparing the sequences of the 29 lineages with the standard bovine DNA sequence (Anderson *et al.*, 1982). Associations were analysed between these polymorphisms and yield traits using a full animal model regressing yield traits on mtDNA genotype based on the polymorphisms identified. The results from this study only provided evidence of negative effects on yield traits of mtDNA mutations. They also performed analysis on cows from North Carolina Herds and found surprisingly high within maternal line sequence variation although the sample size was small. In order for this sort of analysis to be of use it would be necessary to understand better the specific effects of the polymorphisms being analysed.

1.6 Potential Effect on Predicted Breeding Value of Ignoring Maternal Lineage

Initially O'Neill and Van Vleck (1988) addressed the question of the impact of cytoplasmic effects on selection in dairy cattle. They used two approaches to the question, one, what would be the effect of inflated heritability estimates from daughter dam regression techniques and two, what would be the effect of ignoring a cytoplasmic component? Their conclusions were that even if an upward bias occurs in heritability and expected response is overestimated then there is little change in actual genetic gain. Selection using cytoplasmic effects would not increase gain considerably because the only pathway to select for mtDNA is dam to daughter and this is the least intense selection pathway. Boettcher *et al.*, (1996b) addressed the issue of the effect of ignoring cytoplasmic inheritance on the accuracy of predicted breeding values (PBV). They performed a simulation experiment, which generated a population similar to the North American Holstein population but included a component attributable to cytoplasmic inheritance. This simulated data was then analysed using three models. The first model was the same as the USDA model for genetic evaluation of yield traits which has parameters for the following effects; additive genetic, permanent environment, sire x herd interaction, other environmental effects and residual. The second model included maternal lineage as a fixed effect, and the third model treated maternal lineage as a random effect. In order to correctly simulate the data the components of variance applicable to each source in the model needed to be established. Four values of maternal lineage variance were used; 0, 2.5, 5.0, 10% of the variance of random effects. The simulation looked at the consequences of using the different models on genetic response and bull-dam selection. The outcome of this experiment produced results that demonstrated that, even given a 10% variance component attributable to cytoplasmic inheritance, the extra genetic progress from correctly modelling this would only give a 1.75% extra response per year. Various factors contribute to this low response. It was speculated that, using the USDA model without a cytoplasmic component, any maternal lineage effects are partitioned into

PBV and permanent environmental, and therefore indirectly included in the predicted producing ability. In addition to this, cytoplasmic effects have a greater impact on cow PBV than bull PBV, but sire pathways account for more of the genetic response. The greatest effect was seen in the accuracy of selection of dams for sires. These sons are progeny tested across herds, and hence across lineages, and this makes test information relatively robust to cytoplasmic effects. Bull-dam selection was considered in detail and this area appeared to have a strong potential for application of knowledge of cytoplasmic effects. They estimate that with a cytoplasmic effect of 2.5 and 5%, daughters numbers required for bull testing could be reduced by 2 and 4% respectively, which would provide a substantial saving.

1.7 Future Applications

As has been shown, the potential current applications of knowledge of cytoplasmic inheritance would not be large, although there are some circumstances, bull-dam selection for example, that present opportunity. Boettcher *et al.*, (1996c) demonstrated potential improvements to be gained in both genetic response and bull-dam selection given improved knowledge of the underlying genetic model. However in the future there may be more important applications of knowledge of cytoplasmic inheritance. Freeman, (1990), brings attention to the fact that 48 of the proteins involved in the respiratory chain are coded by nuclear genes. We can not therefore consider mtDNA and nuclear DNA in isolation, but should be considering the interaction of the two. This sort of consideration becomes more important in the light of recent advance in the cloning process by nuclear transfer (Wilmut *et al.*, 1997). In this process the actual transfer is of nuclear genetic material. The mitochondria present are those of the animal donating the denucleated ovum (Evans *et al.*, 1999). If cloning is developed as a tool for larger scale applications in animal breeding it will be important to understand the interaction of the two sources of DNA. Indeed cloning or similar molecular work may be useful as a tool in obtaining a better understanding of the effects of cytoplasmic inheritance. Westhusin *et al.* (1996)

have done work on pronuclear transfer procedures, and have also highlighted the potential use of this technique to obtain a better understanding of nuclear/mitochondrial interactions.

Another interesting area for the application of knowledge of cytoplasmic inheritance is in the implementation of MOET (multiple ovulation embryo transfer) schemes. There are several points that need addressing here. MOET provides a means of increasing the genetic progress from the maternal as opposed to paternal route. If superior mtDNA lineages are identified it will be possible to apply this knowledge when selection of donor cows for the scheme occurs. These schemes are based on selection using information from siblings and, for females, the individuals' own performance. The failure to account for cytoplasmic effects in such a scheme will cause biases in the estimation of breeding values with larger effects on bull PBV based upon full and half sister information.

1.8 Conclusions

The question as to what effect mitochondrial DNA has on the inheritance of production traits in dairy populations has been considered using several approaches. However most studies which have estimated a significant maternal lineage variance component for production traits have done so using models such as sire models. These models have been shown to over-estimate maternal lineage variance as they partition additive variance into the maternal lineage component. The implementation of the animal model in more recent investigations has resulted in estimates of less than 1% of phenotypic variance being attributable to maternal lineage variance. A further design improvement would be the use of contemporary records, provided that sufficient maternal family size exists in a contemporary group. This is because in order to isolate a non-nuclear component of variance the optimum strategy is to provide a data structure where animals within a maternal lineage have reduced nuclear genetic relationships. By considering contemporary

animals, where most maternal relationships are no closer than maternal cousins sharing 1/16 of nuclear genetic material, this is achieved.

There is evidence that cytoplasmic inheritance, which is thought to be the cause of maternal lineage variance, does contribute to milk yield persistency (Schnitzenlehner and Essl, 1999) and there are a large number of important traits such as conformation traits that have not been investigated. There has also been very little investigation of the presence of maternal lineage variance in traits such as the efficiency of milk production. Efficiency here is defined as unit milk energy produced per unit feed energy consumed. Such a study is only feasible if individual cow feed intake and milk composition is recorded.

In the following chapters some of the aspects of model design discussed have been implemented to estimate maternal lineage variance. Extensive recording of yield and feed intake is done at the University of Edinburgh Langhill dairy herd. Maternal lineage variance is estimated in Chapter 2 for various traits and derived traits using Langhill information. Chapters 3, 4 and 5 deal with the UK national population. General population parameters are estimated in Chapter 3 using pedigree analysis. Chapter 3 also provides the maternal family structure of the 1997 pedigree registered cows. This information is necessary for the estimation of maternal lineage variance. Chapters 4 and 5 use UK national dairy and conformation records. A contemporary data design is analysed using an animal model in order to estimate maternal lineage variance in both type and production traits. Chapter 6 deals with some conflicts that exist in the estimation of maternal lineage variance in a situation where true cytoplasmic lines exist only when a full pedigree trace is done, coupled with the problems that pedigree errors present in such a trace.

Chapter 2

Estimation of variance of maternal lineage effects at the Langhill dairy herd

2.1 Introduction

Over recent years it has been hypothesised that a component of the phenotypic variance of economically important traits in dairy cattle is inherited exclusively through the maternal line (e.g. Bell *et al.,* 1985; Schutz *et al.,* 1992, Chapter 1). This component of inheritance is of cytoplasmic rather than nuclear origin. It is a long held belief by cattle breeders that some cow families are more important in breeding terms than others, and cows from these families will be used as breeding animals in preference to other cows with similar breeding values. There is also evidence to show that estimates of heritability from daughter dam regressions are higher than those estimates from paternal half-sib analysis (e.g. Seykora and McDaniel, 1983; Visscher and Thompson, 1992). This may suggest that there is a mechanism of inheritance, in addition to nuclear genetic inheritance, that is being transmitted through the female line, which is not being accounted for by current evaluations. A possible explanation for this is the almost exclusively maternal transmission of mitochondrial DNA (mtDNA) in mammalian species (Hutchinson *et al.,* 1974).

Although mtDNA encodes only about 0.05% of all the nuclear plus cytoplasmic genes it is the nature of the potential contribution to phenotypic variation that makes it important. The mitochondria are the energy factories of the cell, and mtDNA codes 12 of the 60 proteins involved in the respiratory chain. For this reason mtDNA may be responsible for a significant component of the phenotypic variation of energy dependent production traits, such as the production of milk.

Early attempts to estimate the magnitude of phenotypic variance attributable to cytoplasmic inheritance treated maternal lineage as a fixed effect in a least squares model of analysis (Bell *et al.,* 1985; Huizinga *et al.,* 1986). These early analyses have been criticised for detecting lineage components that could be explained by sampling variance and residual additive genetic correlations among members of a lineage, not being correctly accounted for by the model applied (Kennedy, 1986). In a more recent study, Boettcher *et al.* (1996b) investigated what the most appropriate method of analysis was and concluded that considering cytoplasmic lineage as a fixed or a random effect in an animal model produced very similar results. Boettcher *et al.* (1996b) simulated data which included a maternal lineage (ML) effect and compared the impact of treating ML as fixed or random in the estimation of breeding values. Their results demonstrated a higher correlation of EBV (estimated breeding value) with TBV (true breeding value) if ML was fitted as a random effect in the model It can be argued that in the commercial situation a large number of small maternal lineages will be encountered, and their treatment as random effects is therefore more appropriate. Further, the mtDNA genomes represented in the study are only a sample from the base population.

The aim of this study was to estimate the magnitude of the maternal lineage component of milk production traits and other measures of efficiency. The Langhill dairy herd is involved in a long-term selection experiment that has been running since 1973. Extensive recording of production traits, feed intake and animal condition has been undertaken, thereby allowing a more in-depth analysis of the traits highly dependent on energetic mechanisms, for which estimates of the ML component has not been previously undertaken. Given the strong link between energy metabolism, feed intake and the efficiency of production these traits may be significantly influenced by a ML component. Due to the nature of the experimental herd, environmental randomisation has occurred and management of all cows is well recorded and can be accounted for, which should eliminate any effects due to preferential treatment of individual cow families. A second aim of the study was to investigate what bias was being introduced into the estimation of a

maternal lineage variance component by using sire half-sib relationships in the relationship matrix of the animal model The sires used at the Langhill herd have been selected within the national population and an anticipated reduction in between sire variance will not be accounted for when using Langhill data, even when an animal model is used for analysis. The bias caused by selection of sires outside of the data set has not been accounted for in previous investigations that have estimated the magnitude of the variance of the maternal lineage component using an animal model (e.g. Boettcher *et al.*, 1996a).

2.2 Materials and Methods

2.2.1 Data Description

1118 records of 517 cows calving between 1982 and 1997, all with a first lactation record, were extracted from the Langhill database.

The traits investigated were all taken from the first 26 weeks of lactation. The methodology of the data collection and the treatment of missing records was outlined in detail by Veerkamp *et al.* (1994). The traits used in the analysis were average daily yield of milk (MLK), fat (FAT), protein (PRT), and average fat and protein percentage (FATP and PRTP). In addition, average daily feed dry matter intake (DMI), net energy of milk production (NE), defined as NE = (0.376*FATP + 0.209*PRTP + 0.9480)*MLK, and a raw measure of efficiency (EFF), defined as EFF = NE / DMI, were analysed. EFF is the average daily energy content of the milk yield (MJ), as calculated by Simm *et al.* (1991), divided by the average DMI (kg) of the cow.

In addition the condition score, on a scale of 1 for lean to 5 for fat, of the cows recorded 48 hours post calving (CC) and the average weekly condition score of the cows over the first 26 weeks of lactation (AC) were also analysed.

All cows were traced back to founder ancestors in the Langhill herd. These founder ancestors were then traced using the UK registration records of the Holstein Friesian Society to either the first point of registration of a founder female or to a cut off point of the year 1920, given that no further convergence looked probable. The earliest cows traced were taken as being the points of cytoplasmic origin for the Langhill herd. This tracing resulted in 517 cows being assigned to 56 cytoplasmic lineages, which define the maternal lineages. The largest maternal lineage contained 72 cows with records within a structure that comprised 49 daughter dam pairs and 38 maternal half-sibs. 9 maternal lineages had only one cow with records. Cows were traced up to 16 generations in the establishment of all the points of cytoplasmic origin in the herd. 13% of the records were allocated to MLs of less than 5 cows with records.

2.2.2 Data Analysis

Data were analysed with REML VCE (Groeneveld, 1996) using a univariate animal model. Three approaches were taken. The first fitted a full animal model with the pedigree fitted including sire, dam, paternal grandsire and granddam, and maternal grandsire and granddam, to give a total of 1212 animals. The second and third approaches were designed to investigate the effect of national sire selection on the variance components, and fitted two alternative animal models, both having a reduced pedigree structure with only daughters, dams and granddams present, a total of 834 animals. In these two models sires were treated as unknown parents in the calculation of A inverse. In the second model, sire was fitted as a fixed effect, and in the third each trait was regressed on the sires EBV for that trait where possible. For the feed and condition traits, where sires

EBVs were not available, the data were regressed on the sire fat yield EBV. All three models were fitted with and without a random component to account for maternal lineage. In summary:

2.2.3 Model

Model 1

$$Y_{ijklmn} = L_i + F_j + YS_k + AL_l + a_m + p_m + c_n + e_{ijklmn}$$

Model 2

$$Y_{ijklmno} = L_i + F_j + YS_k + AL_l + S_o + a_m + p_m + c_n + e_{ijklmno}$$

Model 3

$$Y_{ijklmn} = L_i + F_j + YS_k + AL_l + a_m + p_m + c_n + b_1 \text{ (sirePTA)} + e_{ijklmn}$$

where $Y_{ijklmno}$ = trait; L_i = fixed effect of line, (i=1,2 representing selection or control line); F_j = fixed effect of feed type, (j=1,2 representing high or low concentrate feed); YS_k = fixed effect of year-season of calving, (season was divided into two six month periods

from June to November and December to May); AL_l = fixed effect of age at calving within lactation (l = 1,8); a_m = additive genetic effect of the animal (m=1,517); p_m = permanent environmental effect of animal; c_n = random effect of maternal lineage (n=1, 56); S_o = fixed effect of sire (o=1, 77); b_1(sirePTA) = linear regression of Y on sire predicted transmitting ability ; $e_{ijklmn(o)}$ = residual error. The model was applied to the data with and without the random effect of maternal lineage. The maternal lineage groups were distributed evenly across the selection and control lines and also across the feed types. A subset of the data containing only first lactation records was also analysed using the above models, fitting age at calving as a linear and quadratic covariable. The test statistic used was the likelihood ratio test (LRT) , which is discussed in section 2.4.

In order to investigate whether or not very small lineages were causing confounding between nuclear genetic effects and cytoplasmic effects, the traits that approached a significant component of variance attributable to ML in the first analysis, i.e. FAT, NE and EFF, were reanalysed using two reduced data sets. The first included only maternal lineages that had 5 or more cows with records, which resulted in 448 cows with 973 records in 27 maternal lineages. The second data set had cows with 10 or more cows per lineage, which resulted in 374 cows with 812 records in 16 maternal lineages.

2.3 Results

Overall means and standard deviations of the traits used in the analysis are presented in Table 2.1.

The results of the analyses of production traits are in Tables 2.2, 2.3 and 2.4. Estimates of repeatabilities of all the yield traits, using the full pedigree (Table 2.2), were about 10% lower than recent estimates from data of the UK based Dairy Information system DAISY

Table 2.1. *Trait information for 1118 records and a subset of first lactation records of 517 cows in 56 maternal lineages.*

		First Lactation		All Records	
Trait	Abbreviations	Mean	SD	Mean	SD
Milk (kg/day)	MLK	23	4.8	27	6.4
Fat (kg/day)	FAT	0.96	0.19	1.1	0.25
Protein (kg/day)	PRT	0.73	0.15	0.84	0.20
Fat (%)	FATP	4.2	0.49	4.1	0.51
Protein (%)	PRTP	3.2	0.23	3.2	0.23
Dry Matter	DMI	15	2.1	16	2.8
Efficiency	EFF	5.0	0.77	5.1	0.78
Net Energy	NE	73	14	84	19
Average Condition	AC	2.6	0.31	2.4	0.42
Calving Condition	CC	2.7	0.22	2.6	0.35

(Pryce *et al.,* 1998). The heritability estimates were higher by 9%, 22% and 16% respectively for milk, fat and protein yield, than this study but the heritabilities obtained were very similar to those obtained by Pander *et al.* (1993) in a study of the UK national population. The results of all of the analyses of production traits of the full data set failed to demonstrate any significantly detectable component of variance attributable to ML at the 5% threshold (Tables 2.2, 2.3 and 2.4). This was demonstrated by either the standard errors (from VCE) or taking the log-likelihood ratio test as the test statistic (LRT). Fat

yield gave a LRT very close to significance at the 5% level when analysed using model 3, which fitted a maternal pedigree within the animal model and regressed the data on the sire EBV, with a 4% variance component attributable to ML The inclusion of a ML random component in the model resulted in a repartitioning of the variance components for fat yield, increasing the permanent environment component from 9% to 13%, and reducing the additive variance from 42% to 34% of the total variance. The repeatability estimate for all models did not differ significantly if a maternal lineage component was fitted, indicating that no repartitioning of the residual error variance occurred. Boettcher *et al.* (1996a), performing an analysis using an animal model and treating maternal lineage as random, reported maternal lineage variance components of 0.38, 0.71 and 2.90% respectively for milk yield, fat yield and fat concentration. The data of Schutz *et al.* (1992), when re-analysed using an animal model, as reported by Boettcher *et al.* (1996b), gave maternal lineage variance components of zero for yield traits and 5.6% for the fat concentration trait. This is in contrast to the present study, which found a higher variance component for the fat yield trait and no significant component for fat concentration.

When the energy traits were analysed dry matter intake was not found to have any component of variance attributable to maternal lineage, (Tables 2.2, 2.3, and 2.4). Net energy of milk production did however have a 4% component of variance estimated for maternal lineage, though this was not significant (Table 2.2). The efficiency trait, a combination of both intake and net energy of production, had a ML variance component of 3% of phenotypic variance, but this was not significant at the 5% level. A similar result was obtained for the two different approaches adopted to account for sire effects (Tables 2.3 and 2.4).

Table 2.2. *Heritability (±se) estimates etc. for the repeatability model, model 1,with and without maternal lineage fitted*†

	Without Maternal Lineage		With Maternal Lineage			
TRAIT	h^2	p^2	h^2	p^2	f^2	LRT
MLK	0.42 (0.05)	0.13 (0.04)	0.40 (0.06)	0.14 (0.04)	0.02 (0.02)	0.28
FAT	0.49 (0.05)	0.09 (0.04)	0.42 (0.05)	0.11 (0.05)	0.04 (0.02)	2.44
PRT	0.43 (0.05)	0.11 (0.04)	0.40 (0.06)	0.12 (0.05)	0.02 (0.02)	0.43
FATP	0.65 (0.05)	0.16 (0.05)	0.65 (0.05)	0.16 (0.05)	0.00 (0.00)	0.00
PRTP	0.62 (0.05)	0.12 (0.05)	0.62 (0.06)	0.12 (0.05)	0.00 (0.00)	0.00
DMI	0.37 (0.05)	0.22 (0.05)	0.37 (0.06)	0.22 (0.05)	0.00 (0.00)	0.00
EFF	0.38 (0.05)	0.11 (0.04)	0.33 (0.06)	0.13 (0.04)	0.03 (0.02)	1.18
NE	0.45 (0.05)	0.09 (0.04)	0.39 (0.05)	0.11 (0.04)	0.04 (0.02)	1.88
AC	0.38 (0.05)	0.17 (0.04)	0.38 (0.05)	0.17 (0.04)	0.00 (0.00)	0.00
CC	0.25 (0.04)	0.09 (0.04)	0.25 (0.04)	0.09 (0.04)	0.002 (0.01)	0.06

†See Table 2.1 for abbreviations

h^2 is the heritability

p^2 is the proportion of phenotypic variance attributable to permanent environment

f^2 is the proportion of phenotypic variance attributable to maternal lineage

LRT is twice the difference in log-likelihood between the two models

Table 2.3. *Heritability (±se) estimates for the repeatability model, model 2, with and without maternal lineage fitted. Sire fitted as fixed effect*†

	Without Maternal Lineage		With Maternal Lineage			
TRAIT	h^2	p^2	h^2	p^2	F^2	LRT
MLK	0.24 (0.07)	0.23 (0.07)	0.24 (0.07)	0.23 (0.07)	0.00 (0.00)	0.00
FAT	0.40 (0.06)	0.10 (0.06)	0.33 (0.08)	0.14 (0.07)	0.03 (0.02)	1.30
PRT	0.32 (0.07)	0.14 (0.06)	0.32 (0.06)	0.14 (0.06)	0.00 (0.00)	0.00
FATP	0.61 (0.08)	0.17 (0.07)	0.59 (0.09)	0.18 (0.07)	0.01 (0.03)	0.02
PRTP	0.55 (0.08)	0.13 (0.07)	0.55 (0.08)	0.13 (0.07)	0.00 (0.00)	0.00
DMI	0.42 (0.07)	0.12 (0.06)	0.42 (0.07)	0.12 (0.06)	0.00 (0.00)	0.00
EFF	0.24 (0.06)	0.19 (0.06)	0.20 (0.07)	0.21 (0.06)	0.02 (0.02)	0.56
NE	0.32 (0.07)	0.13 (0.06)	0.29 (0.07)	0.15 (0.07)	0.01 (0.02)	0.30
AC	0.32 (0.06)	0.17 (0.06)	0.31 (0.07)	0.17 (0.06)	0.003 (0.01)	0.03
CC	0.24 (0.05)	0.06 (0.05)	0.22 (0.06)	0.07 (0.05)	0.01 (0.01)	0.60

†See Tables 2.1 and 2.2 for abbreviations

Table 2.4. *Heritability (±se) estimates for the repeatability model, model 3, with and without maternal lineage fitted. Trait regressed on sire PTA*†

	Without Maternal Lineage		With Maternal Lineage			
TRAIT	h^2	p^2	h^2	p^2	F^2	LRT
MLK	0.25 (0.06)	0.23 (0.06)	0.25 (0.07)	0.23 (0.06)	0.00 (0.00)	0.00
FAT	0.42 (0.05)	0.09 (0.05)	0.34 (0.07)	0.13 (0.05)	0.04 (0.02)	2.65
PRT	0.33 (0.06)	0.14 (0.05)	0.30 (0.07)	0.15 (0.06)	0.02 (0.02)	0.34
FATP	0.64 (0.06)	0.14 (0.05)	0.62 (0.08)	0.16 (0.06)	0.01 (0.03)	0.08
PRTP	0.58 (0.07)	0.12 (0.07)	0.58 (0.07)	0.12 (0.06)	0.00 (0.00)	0.00
DMI	0.31 (0.06)	0.26 (0.06)	0.31 (0.06)	0.26 (0.06)	0.00 (0.00)	0.00
EFF	0.27 (0.05)	0.17 (0.05)	0.20 (0.06)	0.20 (0.06)	0.03 (0.02)	1.43
NE	0.38 (0.06)	0.10 (0.05)	0.31 (0.07)	0.13 (0.06)	0.03 (0.02)	1.58
AC	0.34 (0.06)	0.16 (0.05)	0.34 (0.05)	0.16 (0.05)	0.00 (0.00)	0.00
CC	0.21 (0.05)	0.11 (0.05)	0.20 (0.04)	0.11 (0.04)	0.002 (0.01)	0.06

†See Tables 2.1 and 2.2 for abbreviations

No significant ML component was found for condition scores of the cows, whether as a single measure taken 48 hours after calving or as an average weekly score over the 26

weeks of recording and indeed when the data was analysed with the ML component model little or no variance was partitioned into ML.

The results from first lactation records alone are not shown here as they provided similar results. Very low or zero LRT values were obtained from the comparison of models with and without a ML component. The fat yield trait had a 2% component of variance partitioned into ML, but the SD associated with this was greater in magnitude than the component, indicating a similar trend to the full data set. The power of detection of a ML component for this reduced data set was lower than that of the repeatability model for the same cows.

When maternal lineage size was restricted to 5 or more cows per lineage using all lactation records, applying model 1, a significant 4% ML variance component was estimated for fat yield and a 4% component approaching significance was estimated for net energy (Table 2.5). With maternal lineage restricted to 10 or more cows per lineage both the net energy and fat yield traits were found to have a significant component of variance attributable to maternal lineage of 5% and 6% respectively (Table 2.5).

Table 2.5. *Heritability (±se) estimates for the repeatability model, model 1, with and without maternal lineage fitted. Maternal lineage greater than or equal to five or ten cows*†

	Without Maternal Lineage		With Maternal Lineage			
TRAIT	h^2	p^2	h^2	p^2	c^2	LRT
FAT (5)	0.44 (0.05)	0.12 (0.04)	0.37 (0.06)	0.14 (0.05)	0.04 (0.02)	3.27
FAT (10)	0.42 (0.05)	0.10 (0.04)	0.34 (0.06)	0.13 (0.05)	0.06 (0.03)	4.18
EFF (5)	0.36 (0.05)	0.12 (0.04)	0.31 (0.06)	0.14 (0.04)	0.03 (0.02)	1.43
EFF (10)	0.42 (0.05)	0.08 (0.04)	0.36 (0.06)	0.10 (0.05)	0.04 (0.03)	2.09
NE (5)	0.41 (0.05)	0.12 (0.04)	0.35 (0.06)	0.14 (0.05)	0.04 (0.02)	2.58
NE (10)	0.40 (0.05)	0.09 (0.05)	0.33 (0.06)	0.11 (0.05)	0.05 (0.03)	3.09

†See Tables 2.1 and 2.2 for abbreviations

2.4 Discussion

Given the nature of mitochondria it would be expected that highly energy dependent traits such as milk and fat yield would be more likely to have a component of phenotypic variance attributable to ML than, for example, protein yield. Previous investigations have tended to provide evidence to reinforce the hypothesis that if the trait is highly energy dependent then a significant ML component can be detected (Boettcher *et al.*, 1996a). The aim of this study was to investigate further the hypothesis that energy dependent traits have a component of phenotypic variance attributable to ML. This was undertaken by analysing feed intake data, recorded at Langhill, and also combining recorded yield and

intake information into a measure of efficiency. The efficiency trait that was used is a crude measure, since DMI was used instead of MJ energy intake, and the ratio of these varies across years, between treatments, and between early and late lactation. However, the model fitted a feed effect to account for high and low concentrate feeding, and also a year-season effect, which accounted in some way for between year variation in feed quality. These effects are expected to account for most of the dietary energy variability.

The results of the investigation of net energy, efficiency and dry matter intake traits failed to provide evidence to support the energetic relationship hypothesis using the full data set. For the intake trait all three models failed to detect a component of phenotypic variance attributable to ML. The LRT for these models were zero. Under the null hypothesis of no variation due to maternal lineage, the asymptotic distribution of the log likelihood ratio test is $1/2\chi^2(0) + 1/2\chi^2(1)$, giving a 5% significance threshold of 2.7 (e.g., Stram and Lee, 1994). It may be the case that the power is too low to detect a small ML effect. In order to investigate the power of detecting a component of variance of the magnitude expected due to maternal lineage a one-way random effect ANOVA was considered. Given a balanced data structure of 500 cows, a similar magnitude to the data set used in this analysis, the power of detection of various maternal lineage variance components was investigated using the F ratio power test (e.g., Lynch and Walsh (1998)). It was shown that 3%, 4% and 5% variance components would be detected with, respectively, powers of 25-70%, 30-80% and 50-90%. This test places an upper limit on the detection of a 3% maternal lineage variance component of 70% power which re-enforces the suggestion that the power of this analysis was too low to detect a small maternal lineage variance component. The energy trait however approached significance with a component of about 4% in magnitude. Previous analyses by Schutz *et al.* (1992) demonstrated a significant effect of maternal lineage on milk energy concentration treating lineage as a fixed effect, but in the present study yield traits were found to show a stronger relationship with maternal lineage than energy concentration traits.

When the data were edited to include only maternal lineages of greater than, or equal to, either 5 or 10 cows per lineage, the LRT was significant at the 5% level for fat yield and net energy of milk production. A possible explanation is that with few cows with records in a lineage the maternal lineage variance is confounded with the additive nuclear variance. Given the low number of records available, the small lineages were used in the analysis to provide a better estimate of the additive variance. The loss in power that would appear to be the result of the inclusion of these lineages was not expected.

The three different models were fitted to quantify, in some way, what effect selected sires were having on the detection of a ML component. The reduction in variance due to sire selection is not accounted for when the full animal model is fitted because not all of the data on which selection was based is present in the analysis. However in treating sires separately from the pedigree, heterogeneous variance between sire and dam estimates is removed from the animal model. By regressing the data on the sire EBV the sire variance is accounted for, if the EBVs are estimated accurately and only a single degree of freedom is lost. Information is lost from the estimation of additive nuclear variance. The theoretical expectation of the estimate of heritability using an animal model with only maternal relationships and removing the variation due to sires is of the magnitude $h^2/(1 - h^2/4)$, where h^2 is the heritability estimate of the full pedigree animal model This represents the loss of residual variance due to fitting the sires as a fixed effect in the model. However it was found that the heritabilities obtained from this model were lower than expected (Tables 2.3 and 2.4). When model 1 was run accounting only for paternal relationships or maternal relationships in the relationship matrix, a higher estimate was obtained for heritability from the paternal model. For milk yield this was 0.62 compared to a maternal model estimate of 0.43 (results not shown in Tables). This difference in heritability estimates was caused by a higher estimate of additive variance from the paternal relationship model. This may be explained by a genetic trend inflating the variance of a

group of selected sires used over a period of fifteen years. This trend was not accounted for in the models. It is apparent from the results that treating sire as a fixed effect and excluding sires from the pedigree (Table 2.3) did not improve the detection of a maternal lineage component. By regressing the data on sire EBV (Table 2.4) the significance of the detection of a component of variance attributable to ML for fat yield increased to a level that was very nearly significant but the magnitude of the component was unchanged from the full pedigree estimate. The regression approach was expected to be superior, as only one degree of freedom was lost. Indeed the results indicated this to be the case. Fitting a linear regression on sire PTA in the model actually increased the LRT for fat yield but still failed to provide a criterion for rejection of the null hypothesis at the 5% level of significance. Given the magnitude of the components of ML variance detected it is difficult to draw any conclusions as to whether or not this alternative treatment of sires in the model provides a better method of estimation.

Another concern was the use of data collected over 15 years during which time the mean milk yield has increased leading to an associated change in variance. Roughsedge *et al.* (1998) applied log transformations to yield data in order to correct for the heterogeneous variance over time due to an increase in mean yield. The analysis of this transformed data provided similar results to the untransformed data and it was concluded that analysing records that were not contemporaries did not adversely affect the results.

Though the analysis of one herd provided limited numbers of records, the maternal pedigree used to assign ML was extensive, in that it provided all known links since 1920. This structure ranged from grade up cows, with only daughters present in the herd, to contemporary animals with 16 generations to their cytoplasmic origin. This assignment of ML is necessary with a small data set but is vulnerable to a reduction in power due to pedigree errors. The effect of pedigree errors occurring in the assignment of lineage has the effect of downward biasing the estimation of a ML component. Gibson *et al.* (1997)

proposed that, given knowledge of the pedigree structure and the pedigree error rate, a correction factor could be calculated to account for the bias. Given knowledge of the error rate and of the information obtained by tracing further generations in the maternal lineage, an optimum number of generations could be calculated for the assignment of maternal lineage.

Given the detection of a component of variance for maternal lineage in production traits then several implications can be considered. Boettcher *et al.* (1996b) using simulated data investigated the impact of such a component on both the ranking of animals and the selection of bull-dams for progeny testing schemes. They found that there was a greater impact on ranking of cows than bulls but the greatest immediate impact would be on bull-dam selection. Of potentially greater importance is the identification of superior individuals in MOET nucleus schemes where information of full and half sibs is used in the evaluation of donor cows and nucleus bulls. Looking to the future the process of cloning by nuclear transfer (Wilmut *et al.*, 1997) will require the identification of superior cytoplasmic lines as the recipients of nuclear material.

In order to increase the power of estimation for maternal lineage variance, analysis of larger data sets of contemporary records is necessary, although these analyses will be restricted to milk yield traits. Such an analysis on a national UK data set is undertaken in Chapter 5.

Chapter 3

Quantifying genetic contributions to a dairy cattle population using pedigree analysis

3.1 Introduction

Over the last twenty years there has been a rapid increase in the proportion of North American Holstein genes in the British black and white dairy cattle population. More recently genetics from European countries such as France and the Netherlands have been imported. However, these countries themselves have sourced most of their genetics from North America, so that we do not distinguish between North American and European genes in this study. This is not the first time that an importation has occurred in the UK dairy population. Robertson and Asker (1951) provided a brief history of the formation of the British Friesian herd tracing the effects that a variety of different importations had had. They document importations that occurred from Holland in the last century and the early part of this century and also a Canadian importation in 1946. The recent importation has certainly been the most significant due to the advances that have occurred in reproductive technology allowing a more rapid and widespread grading up of animals. Within this study the British Friesian population can be defined as those individuals registered with a British Friesian breed code in the database and foreign animals can be defined as other breed codes, those individuals being mostly of North American Holstein origin. The breed substitution has had an effect on the degree of relationship and level of inbreeding in the British dairy population, and the extent of this was investigated. Inbreeding trend has for many years been the preferred tool for the quantification of the rate of genetic drift but, as highlighted by Boichard *et al.* (1997), in order for the measure to be of relevance the population should conform to the rules specified by Wright (1931). It should be closed, unselected, panmictic and of finite size. It is quite obvious that the British dairy population does not satisfy these requirements. Boichard *et al.* (1997) proposed the use of measures of genetic variability more commonly used in conservation

genetics. Lacy (1989) developed a 'founder equivalent' approach which combined the information of the founder animals contributing to the population under study and estimated the number of equally contributing ancestors that would provide the same level of genetic diversity. Caballero (personal communication) indicated that this parameter is directly related to group coancestry (Cockerham, 1967), the average pairwise coancestry of a group of individuals in a pedigree including reciprocals and self-coancestries. The parameter can also be related to genetic diversity defined as the expected frequency of heterozygotes by descent. This approach was taken further by Boichard *et al.* (1997) in the development of 'founder ancestor equivalent' which located the highest contributing ancestors to the population under study and used these animals to estimate the equivalent number of equally contributing ancestors that would provide the same level of genetic diversity. Bowman *et al.* (1978) studied the degree of relationship and average level of inbreeding in the British Friesian population. This current study is intended to update the information on the British dairy population and to investigate other properties of the dairy population by estimating parameters of genetic diversity.

3.2 Material and Methods

3.2.1 Data

Data were extracted from the database of the Holstein Friesian Society of Great Britain and Ireland which has complete pedigree records, including grade-up animals, back to 1960. In this study grade-up animals are the progeny of non-pedigree cows that have been mated to pedigree bulls. This database also records the pedigree of all imported foreign cows and, if available, three generations of ancestors. In total the database holds about 5 million animals. Random samples, each of two thousand cows, were taken from cows born in specific years and the complete ancestry of these animals present in the database was traced. This pedigree file was then used for the various analyses. Ten samples were

taken for 1997 births, which was essentially the year of interest. A further 4 samples per year were then taken going back in 5 year intervals including 1992, 1987, 1982, 1977, 1972, 1967. Five years was taken as being approximately the equivalent of a generation. The use of sampling rather than analysing the whole population was necessitated by a restriction on computing resources.

3.2.2 Founder equivalent

Founder animals were taken as being ancestors having unknown parents. If only one parent was unknown the animal was treated as a half founder and its contribution to the population under consideration was halved. The formula for the calculation of the founder equivalent (FE) parameter is:

$$FE = 1/\sum_{k=1}^{f} q_k^2$$

where q is the proportional contribution of a founder k (total number of founders is f), to the generation for which FE is being calculated. If all founders had contributed equally to the population then the FE would equal the actual number of founders. This provides a measure of the change in genetic diversity of the population over time.

3.2.3 Founder ancestor number

The FE parameter does not properly account for bottlenecks in the pedigree caused by the heavy use of AI in cattle populations. If a recent sire made a high contribution to the current population through AI then it would be more important, in respect to the loss of

genetic diversity, than tracing all of the pathways that founder ancestors in the pedigree made through this individual In order to account for this Boichard *et al.* (1997) proposed a second measure, founder ancestor number (FA), which treats all the ancestors in the same way regardless of whether they are founders. The highest contributing ancestor, termed a 'pseudo-founder' was identified, its contribution stored, and then it was removed from the pedigree so that no further contribution could be made through its pathway in the pedigree. The procedure was iterated, finding the highest marginal contribution at each cycle. At each round of iteration an upper and lower limit was imposed on FA. The upper limit (f_u) was determined by setting the remaining contribution to the population genome to be equally distributed across all remaining founders and calculating FA. The lower limit (f_l) was determined by considering that all remaining founders contribute the proportion of the genome equivalent to the last found pseudo founder. The upper and lower limits converge over iterations and a criterion for stopping can be imposed on this process using either a level of convergence or a given number of pseudo founders. In this study a stopping criteria of f_u-f_l<5 was used.

3.2.4 Founder origin

The contribution of the founder animals to the average individual genome of the population was determined in order to show the influence that the importation of the North American genome has had over the last thirty years. This involved finding the earliest known ancestors of the cows in the sample and calculating their contribution to the current gene pool based on each parent contributing 50% of their offspring's genome. The information generated by the FE procedure was used for this analysis.

3.2.5 Inbreeding coefficient and average relationship

Inbreeding coefficient was determined using the methodology of Meuwissen and Luo (1992). This was based on the decomposition of the additive genetic relationship matrix

A, as described by Henderson (1976): **A** = **LDL'**, where **L** is a lower triangular matrix containing the fraction of genes that animals derive from their ancestors, and **D** is a diagonal matrix containing the within family additive genetic variances of animals. **L** is calculated row by row in the algorithm. The results of the samples for a given year were pooled. Average coefficient of inbreeding was determined for the whole population and also for those individuals with a non-zero coefficient of inbreeding. The average relationship between all members of a given sample, born in the year of the sample, was determined using the recursive algorithm of Miglior *et al.* (1992).

3.2.6 Maternal family distribution

In addition to the nuclear genome the current distribution of the mitochondrial genome in the Holstein Friesian cow population was analysed using all cows born in 1997 as the study population. The earliest maternal ancestor for each of the Holstein Friesian cows born in 1997 was traced using the same database as the other analyses and this was taken as being the point of cytoplasmic origin. Mitochondrial DNA (mtDNA) is inherited almost exclusively maternally in mammalian species (Hutchinson *et al.*, 1974) and as such all cows in a maternal family will have the same mtDNA as the cow that provides the cytoplasmic origin of that maternal family.

3.3 Results

3.3.1 Samples

For all parameters estimated the samples provided consistent results within the year sampled. For example in 1997 the mean and standard deviation of the sample results was 0.43±0.057% for average inbreeding coefficient, 1.34±0.09% for average degree of relationship and 93±4.26 for FE. The results presented are a mean of the sample results

for each year. Table 3.1 gives details of the sample pedigrees and shows that the 'depth' of pedigree information changed in the early 1990's indicative of the change to North American Holstein genome origin, which has less ancestral information stored on the database.

Table 3.1. *Average information about pedigrees samples[a] of 2000 cows*

Year of birth	Number of founders in sample pedigree	Number of ancestors in sample pedigree	Average number of ancestors per individual
1967	2535	2785	2
1972	3523	7492	53
1977	4349	11842	254
1982	8635	33210	872
1987	8414	31361	1321
1992	6090	22284	796
1997	6640[b]	26434[c]	646

[a]10 samples in 1997, 4 samples in other years

[b]SD between 1997 samples134

[c]SD between 1997 samples 454

3.3.2 Founder Origin

The change in the origin of the average genome over time (Table 3.2) clearly demonstrates the influence of the North American Holstein breed on the British Friesian population. The change in origin was most clearly seen over the last ten years when male Holstein origin rose to the same level of influence as British Friesian females had had

twenty-five years ago. Holstein female origin had risen to 20% in 1997 but the rate of increase was seen to fall to only 2% (0.4%/year) between 1992 and 1997. In all four origins, except for the Holstein female, the rate of change did not appear to have stabilised by 1997. It is important to note that the completeness of the pedigree information has an influence on the male:female contribution ratio. For example consider a simple scenario of a cow with pedigree information available on the sire and dam and the paternal grandsire and granddam only. If base animals are taken as being animals with unknown parents then 75% (50% dam, 25% paternal granddam) of the genome is traced to female origin and 25% (25% paternal grandsire) to male origin.

Table 3.2. *Origin of nuclear genome of the British Holstein Friesian population*

	British		Foreign	
Year of birth	Female (%)	Male (%)	Female (%)	Male (%)
1967	72	24	2	2
1972	55	37	4	4
1977	53	34	6	7
1982	48	31	9	12
1987	36	28	14	22
1992	22	20	18	40
1997	13	11	20	56

3.3.3 Average Inbreeding Coefficient

The base population of the pedigree samples used in this analysis was 1960. The samples from 1967 were, therefore, not very deep in pedigree structure (Table 3.1) and, as can be seen in Figure 3.2 and Table 3.3, calculated inbreeding had risen little above zero at this

time. Bowman *et al.* (1978) calculated average inbreeding coefficient of cows in the British Friesian population between 1955 and 1972, tracing animals back to the 1914 herd book. They used the sampling techniques described by Wright and McPhee (1925), that involve using three generations of complete pedigree information and then a sampling technique to establish common ancestors prior to these generations. In 1960, the base year in this current study, they estimated an average coefficient of inbreeding of 2.28%. The rate of increase in average coefficient of inbreeding in the study of Bowman *et al.* (1978) between 1965 and 1972 was very similar to that seen in the current study in which the level of inbreeding in the population followed a fairly steady increase between 1967 and 1982, over which time the origins of the population were in a steady state (Table 3.2). However between 1982 and 1987 a steady fall was seen in the average inbreeding coefficient and this fall continued until 1992. From 1992 to 1997 a slight rise of 0.05% was seen, though when the years between 1992 and 1997 were analysed the samples were seen to differ little from 0.4%. More interesting were the results of the non-zero average inbreeding coefficients and the distribution of the inbred individuals. 1967 can be discounted as being too close to the base population to have had a significant level of inbreeding with only 1% of the population inbred, but by 1972 24% of the population was inbred and this rose to nearly 50% in 1977 and 1982. By 1987 the proportion of inbred animals had fallen to 42% and this proportion continued to fall until 1997. It was evident when the distribution of the inbred individuals was examined that the 1997 results were very similar to the 1972 results with only the proportion of individuals more than 12% and less than 1% inbred being lower in 1997. This similarity is not surprising given that essentially the base population of the 1997 animals had changed, as demonstrated in Table 3.2, to the imported Holstein population, a population with less depth of pedigree available. The distribution of non-zero inbreeding coefficients across all years appears to be an exponential distribution with a high frequency close to zero and a long thin tail up to 25%.

Table 3.3. *Average inbreeding coefficient (%) of the British Holstein Friesian population*[a]

Year of birth	Average inbreeding coefficient (%)	SD average inbreeding coefficient	non-zero avg. inb. coeff. (%)	SD non-zero avg. inb. Coeff.	Proportion inbred (%)
1967	0.05	1.22	4.05	3.87	1
1972	0.55	2.16	2.33	3.97	24
1977	0.82	2.54	1.75	3.46	47
1982	0.74	2.31	1.55	3.15	48
1987	0.52	1.90	1.17	2.76	42
1992	0.38	1.70	1.60	3.24	23
1997	0.43	1.67	2.32	3.24	19

[a] Base population 1960

3.3.4 Average Degree of Relationship

The average degree of relationship (Figure 3.3, Table 3.4) has been less affected by the breed grade-up than the average inbreeding coefficient. Though the parameter fell in 1987 the drop was soon halted with a stable average degree of relationship between 1987 and 1992. After 1992 the level was seen to increase at a rate close to the magnitude of the rate prior to 1982. However, what is more interesting is that the proportion of inbred individuals fell in 1987 and 1992 but by 1997 the figure had risen, with 93% of the population showing a non zero degree of relationship.

Table 3.4. *Average pair-wise relationship of the British Holstein Friesian population*

Year of Birth	avg. rel. (%)	SD avg. rel.	non-zero avg. rel. (%)	SD non-zero avg. rel.	non-zero proportion (%)
1967	0.16	3.94	1.83	1.17	9
1972	0.52	2.62	1.08	1.90	48
1977	0.90	2.46	1.14	2.22	79
1982	1.25	2.64	1.46	2.50	85
1987	1.02	2.47	1.28	2.26	80
1992	1.03	2.94	1.52	2.50	60
1997	1.34	2.87	1.80	2.59	93

3.3.5 Founder and ancestor equivalent number

The FE and FA parameter estimates for the samples (Table 3.5) illustrate the fall in genetic diversity over the last thirty years. The values shown for FA were actually the upper limits given the convergence criterion that $f_u - f_l >= 5$, and therefore potentially over-estimated by 5. The convergence of the upper and lower parameters is illustrated in Fig. 3.1. The trend in FE and FA shows that the fall in genetic diversity of the population was greatest in the years close to the base population.

Table 3.5. *Parameters of genetic diversity of the British Holstein Friesian population*[a,b]

Year of birth	Number of founders	FE	FA	FE as a percentage of founders (%)	FA as a percentage of founders (%)
1967	2535	1353	702	53	28
1972	3523	586	276	17	8
1977	4349	337	175	8	4
1982	8635	237	169	3	2
1987	8414	227	163	3	2
1992	6090	151	144	2	2
1997	6640	93	93	1	1

[a] FA is the founder ancestor equivalent number

[b] FE is the founder equivalent number

From 1977 to 1987 the FE parameter fell very little and the FA parameter, which estimated variability based on the highest contributing ancestors remained constant from 1977 until 1992. In 1997 both parameters had fallen to the same value of about 90. The convergence of these parameters in 1997 is indicative of the change in the founder population to the imported Holstein population with few selected sires being used to grade the population up. These few, but highly influential, founder individuals were also identified as being the highest contributing ancestors and hence the two parameters were seen to converge. To further illustrate the effect of few highly selected sires being extensively used, Table 3.6 shows the highest ancestor contributions to the years investigated. The highest contributing ancestor in 1997, To-Mar Blackstar, was responsible for 5% of the average nuclear genome in the population, more than double the level seen in 1967. Hanoverhill Starbuck made the second highest contribution, of 4.6%

of the nuclear genome, in 1997, three times the contribution of the second highest contributing ancestor in 1967. The top 50 ancestors in 1997 were responsible for 50% of the nuclear genome, more than 3 times the 1967 level

Table 3.6. *Most important ancestor contributions*

	1967	1972	1977	1982	1987	1992	1997
First ancestor †	0.023	0.032	0.033	0.026	0.028	0.041	0.051
Second ancestor †	0.017	0.028	0.030	0.025	0.022	0.034	0.046
Third ancestor †	0.007	0.013	0.023	0.024	0.021	0.025	0.030
First 10 ancestors †	0.080	0.140	0.165	0.201	0.185	0.197	0.282
First 50 ancestors †	0.165	0.301	0.359	0.409	0.440	0.427	0.501

† Proportion of alleles contributed by .....

3.3.6 Maternal family distribution

The distribution of cows across maternal families can be seen in Table 3.7. Most of the cows are seen in families of less than 5 cows. The average family size was 2 with a standard deviation of 2.75 and nearly 70% of the families had only one cow. These single cow families represent 32% of the registered cows born in 1997 and traced on average 3 generations to their cytoplasmic origin, with about 70% tracing to grade-up cows. When the range of family size was increased to families of 1-4 cows 93% of the families were accounted for containing 71% of the cows. Only 7 families were greater than 100 cows, with one having 217 members.

Table 3.7. *Distribution of Holstein Friesian cows born in 1997, across maternal families*

Family Size (no. cows)	No. Families	No. Cows	Proportion of cows (%)	Average no. of Generations to Origin
1	69575	69575	32.1	3.3
2-5	36009	96212	44.4	4.7
6-10	3743	27277	12.6	6.0
11-20	1017	13986	6.5	6.4
21-30	150	3721	1.6	6.8
31-40	56	1932	0.9	6.8
41-100	49	3046	1.4	7.4
101-150	5	574	0.3	7.7
151-220	2	383	0.2	9.0
Total	110607	216706	100	7.6

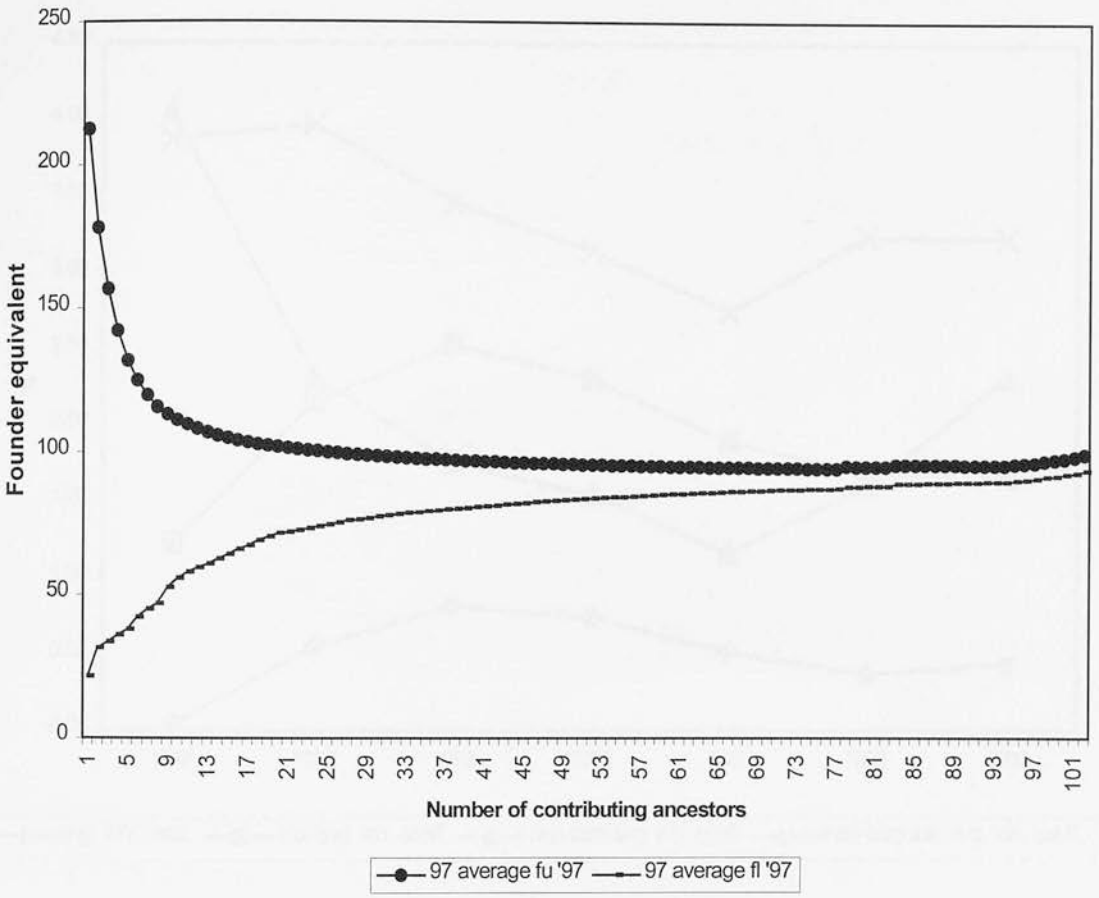

Figure 3.1 Convergence of theoretical upper and lower limit of founder equivalent number over highest contributing ancestors

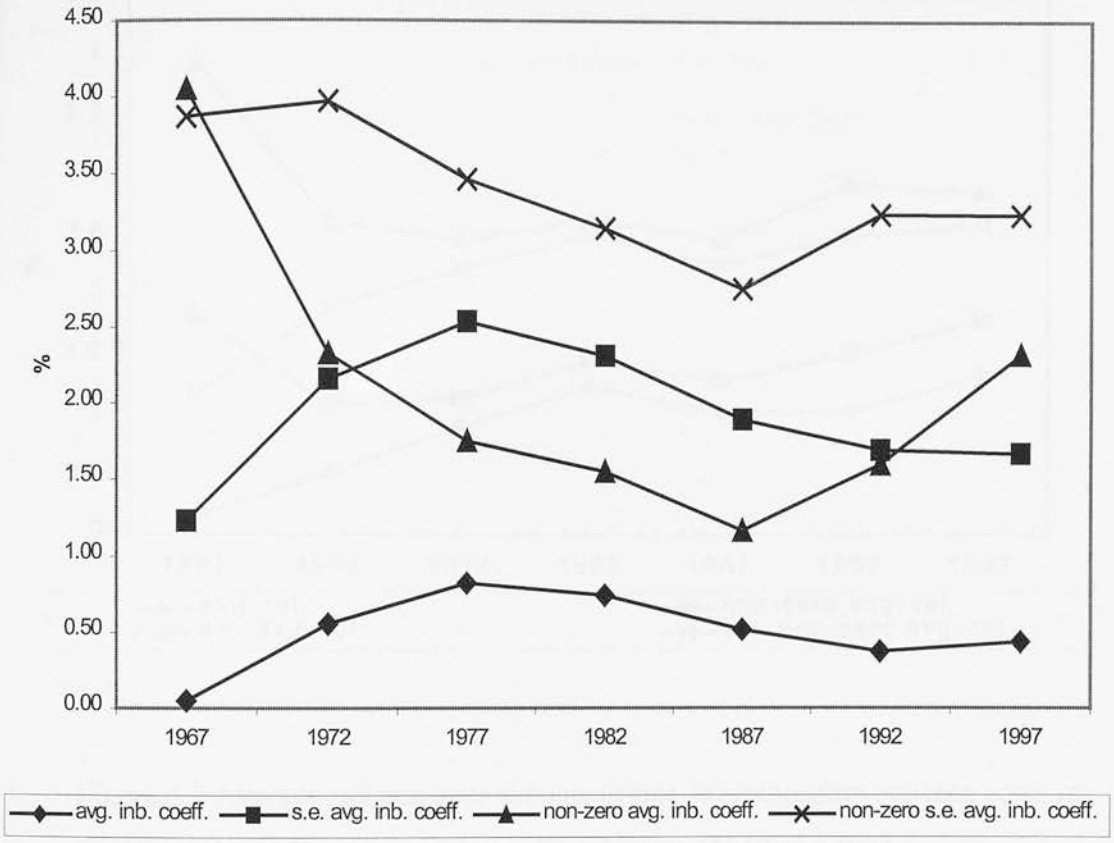

Figure 3.2 Average and non-zero average inbreeding coefficient of cows of the UK dairy population born between 1967-1997

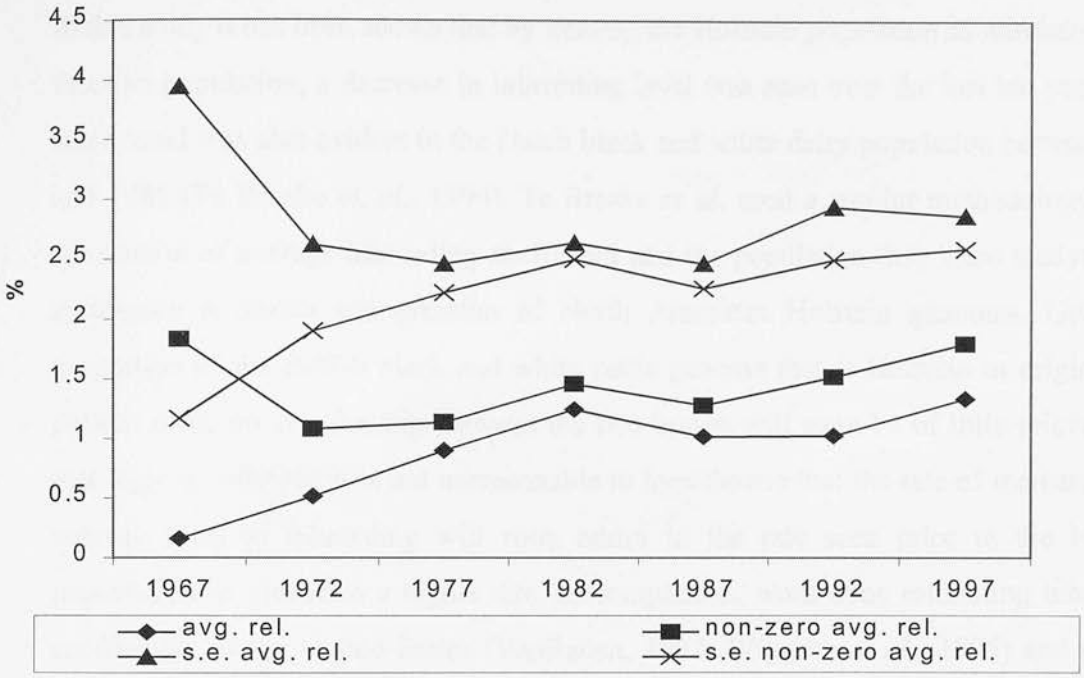

Figure 3.3 Average and non-zero average degree of relationship between cows in the UK Holstein Friesian population born between 1967 and 1997

3.4 Discussion

In this study it has been shown that by treating the Holstein population as unrelated to the Friesian population, a decrease in inbreeding level was seen over the last ten years. The same trend was also evident in the Dutch black and white dairy population between 1980 and 1985 (Te Braake *et. al.*, 1994). Te Braake *et al.* used a similar methodology in the calculation of average inbreeding coefficient and the population they were studying had undergone a similar introgression of North American Holstein genomes. Given the proportion of the British black and white cattle genome that is Holstein in origin at the present time, the relationship between the two breeds will soon be of little relevance. If this logic is followed it is not unreasonable to hypothesise that the rate of increase in the average level of inbreeding will soon return to the rate seen prior to the Holstein importation or indeed to a higher rate. In comparison, work done estimating inbreeding coefficients in the United States (VanRaden, 1992; Wiggans *et al.*, 1995) and Canada (Miglior and Burnside, 1995) has consistently shown higher levels of inbreeding than has been seen in the British Holstein Friesian population. VanRaden (1992) and Wiggans *et al.* (1995) using the same population of US Holsteins with a 1960 base population calculated the average inbreeding coefficients as 0.4%, 1%, 2% and 2.6% respectively for 1970, 1980, 1987 and 1990. Miglior and Burnside (1995) using a base population of 1950 estimated an average inbreeding coefficient of 1.7% between 1986 and 1990. They showed a plateau from the late 1970's until the late 1980's due to the heavy use of US Holstein sires in the Canadian population.

The average degree of relationship parameter would appear to have been less sensitive to the introduction of Holstein genes than the level of inbreeding. Indeed given the steady increase in average relationship and the introgression of North American Holsteins across most of Europe, it is inevitable that the average inbreeding coefficient will increase in the future. In the study of Bowman *et al.* (1978) the degree of relationship estimated for 1960

was 1.45% but was seen to have risen to 3.19% by 1972, with the rate of increase between 1965 and 1972 almost identical to that estimated by this current study.

As has previously been discussed the recent change in breed origin resulted in both the rate of increase in average inbreeding coefficient and average degree of relationship to be lowered considerably and indeed a decrease in the inbreeding parameter was seen during the late 1980's and early 1990's. These parameters are not therefore able to provide a good estimate of the genetic diversity of the British black and white dairy population over recent years. FE and FA were used to illustrate the change in genetic diversity over time, whilst the population was undergoing a major grade-up, which was not possible using the average inbreeding coefficient. These parameters provided a useful means of describing the change in genetic diversity but as indicated by Boichard *et al.* (1997) they cannot be used as predictors of future variability. In previous studies using FE and FA (Boichard *et al.*, 1997; Sölkner *et al.*, 1998) whole populations have been used. In this study samples from the population were used and whilst the parameters provided a description of the change over time they did not necessarily accurately estimate the magnitude of the population parameters. However, given the nature of these parameters, i.e. that they are heavily dependent on the proportional contribution of the top few, highest contributing, founders/ancestors it can be argued that the estimates provide a close approximation. The identity and magnitude of the contribution of the highest contributing ancestors was found to be consistent across the samples for given years, and this further validated the accuracy of the sample result. Young and Seykora (1996) estimated the ancestors with the highest contribution to the 1990 US Holstein females. Two of the three bulls with the highest contribution to the US female genome in 1990 were the same two bulls that made the highest contribution to the British Holstein Friesian female genome in 1992. Round Oak Rag Apple Elevation contributed 12.2% and 3.4% and SWD Valiant contributed 9.6% and 4.1% to the US and UK female populations respectively.

The use of long-term founder contributions is similar to the approach taken by Woolliams and Mantysaari (1995) in the Finnish Ayrshire population, however they related this contribution to the rate of inbreeding using the methodology of Wray and Thompson (1990).

The use of founder equivalent as a measure of genetic diversity can be criticised due to the nature of long-term founder contributions. Over a number of generations the proportionate contributions of founder individuals will stabilise to the same across all contemporary individuals (Wray *et al.*, 1994) and hence the founder equivalent parameter will become fixed. Caballero (personal communication) states that minimising coancestries is a more effective way of maintaining genetic variability than trying to equalise founder contributions. This minimises the variances of contributions from ancestors to descendants in all previous generations to the current one, *i.e.* maximising the effective population size.

Rate of increase in average inbreeding coefficient has been used frequently as a measure of genetic diversity (e.g. Goddard, 1992; Wang, 1997). However it is a measure that is sensitive to changes in population structure and crossbreeding. One method to avoid this problem was proposed by VanRaden (1992). He suggested that given that Holstein and Friesian populations will have a common base at some point in the past, then it is possible to assign an average relationship between Friesian and Holstein individuals based on several assumptions utilising earliest known ancestor information. The genetic diversity parameter FA estimated in this study is less sensitive to a change in origin and reflects the level of contribution being made by few very influential individuals to the whole population. Such information could also be used to identify families that have high influence and provide extra information to be used in future selection strategies. These measures of founder representation are used by conservation biologists in order to retain genetic variation of captivity bred 'wild' populations (Lacy, 1989). As FE and FA

decrease so the level of inbreeding will increase with the general increase in similarity of the genome. The present study estimated the genetic diversity parameters to be of very similar magnitude to those estimated by Boichard *et al.* (1997) and Sölkner *et al.* (1998) for dairy breeds. However the FA and FE parameters converged in 1997 in the British Holstein-Friesian population. This was attributable to the fact that the imported Holstein sires formed the new founder base population and, at the same time, these few, heavily used sires made a large contribution to the average genome of the population. This again illustrates the robustness of the FA parameter, which does not rely on distant relationships to the same level that the other parameters do. The measure of FE in small conservation populations is used as a standard by which genetic diversity can be maintained. Here the populations are not being actively selected and the objective is to attain an equal representation of all founders across generations. In livestock improvement the objective is very different but if we are looking towards long term effective improvement then preservation of the variability of the genetic base is important. Goddard (1992), considering the global black and white population and using N_e as the measure of genetic diversity calculated the optimal effective size using discounting to estimate future gains. Such a calculation is dependent on several assumptions regarding objectives and differences between countries. Goddard (1992) hypothesised that differences between countries in both environment and objectives will lead to the development of isolated strains, which originate from the same founding animals, i.e. North American Holsteins. What is not certain at present is the degree of exchange of genes that will continue to occur between these 'isolated' populations.

When the mitochondrial genome was considered, using 1960 as the population of cytoplasmic origin, a very different level of diversity was seen. Due to the nature of the transmission of the mitochondrial genome, i.e. it is almost exclusively maternally inherited, the heavy use of foreign sires has had no impact on its diversity. This study provided an underestimate of maternal family size for two reasons. Firstly, pedigree information prior to 1960 was not used and secondly many of the single cow families

were from a first generation grade-up cow with no information on the non-pedigree dam. It would appear that the level of diversity remaining between mitochondrial genomes is still high or at least has not been reduced greatly by selection over the last thirty years.

Chapter 4

Effects of cow families on type traits in dairy cattle

4.1 Introduction

Cattle breeders have long held the belief that certain maternal lineages, referred to as cow families, have special attributes. Gibson *et al.* (1997) suggested that it is not uncommon for dairy sire analysts to require that a potential bull-dam has not only a high genetic evaluation for type, but should also come from a cow family with several successive generations of outstanding females. Indeed several researchers have found higher estimates of heritability for production traits from daughter-dam regression than from paternal half-sib correlation (e.g. Seykora and McDaniel, 1983; Visscher and Thompson, 1992). This may suggest that there is a component of phenotypic variance that is being exclusively maternally inherited in addition to additive genetic inheritance, which can be attributed to cow family.

Preferential treatment of cows from cow families that are thought to be of outstanding merit could be a potential source of bias in the estimation of cow family variance. It is possible to envisage that extra management given to cows from such lineages from birth could be responsible for a component of variance attributable to cow families. This influence would however be expressed through production traits and it is less obvious what the impact of preferential treatment would have on type traits. It is possible that better veterinary care and better presentation of cows at the time of type assessment could have an effect on known cow families.

Mitochondrial inheritance has been hypothesised as a possible mechanism to account for the cow family effect. In mammals the mitochondria is almost exclusively maternally

inherited (Hutchinson *et al.*, 1974) and contains a closed loop of 16,338bp of DNA, coding 12 of the genes involved in energy production via the electron transport chain (Anderson *et al.*, 1982). Over recent years several studies have estimated the magnitude of the maternal lineage component of phenotypic variance responsible for production traits in dairy cattle (e.g. Boettcher *et al.*, 1996; Schutz *et al.*, 1992; Chapter 2). The aim of this study is to estimate the size of the maternal lineage variance component responsible for linear type traits and various composite traits in the type classified pedigree UK Holstein Friesian cow population.

4.2 Material and methods

4.2.1 Data description

A total of 33,325 first lactation Holstein Friesian pedigree cows were used in the analysis. All cows were type classified between 1996 and 1998 in the UK.

Details of the type classification scheme operated by the Holstein UK and Ireland (HUKI), formerly the Holstein Friesian Society of Great Britain and Ireland, were given previously by Meyer *et al.*, (1987) and Brotherstone and Hill (1991).

The traits investigated were 16 linear type traits, two farmer scored traits of ease of milking and temperament, four subjective composite traits and an overall type score. The latter trait comprised of the four composites with respective weights of 2(dairy) + 2(body) + 2(legs and feet) + 4(mammary). A summary of these traits can be seen in Table 4.1.

Table 4.1. *Trait information for 33325 type records*

Trait	Abbreviation	Score			
		1	9	Mean	SD
Total Score	TSC			74.18	5.54
Stature	STA	Small	Tall	6.45	1.38
Chest width	CW	Narrow	Wide	5.32	1.48
Body depth	BD	Shallow	Deep	6.11	1.33
Angularity	ANG	Coarse	Angular	5.81	1.39
Rump angle	RA	High pins	Low pins	4.00	1.31
Rump width	RW	Narrow	Wide	5.43	1.47
Rear legs side	RLS	Posty	Sickled	5.50	1.35
Foot angle	FA	Low	Steep	5.28	1.32
Fore-udder attachment	FUA	Loose	Tight	5.22	1.47
Udder support	US	Broken	Strong	5.76	1.42
Udder depth	UD	Below hock	Above hock	5.77	1.37
Teat placement rear	TPR	Wide	Close	4.69	1.42
Teat placement side	TPS	Close	Apart	5.64	1.31
Teat length	TL	Short	Long	4.42	1.40
Condition Score	CS			4.76	1.58
Rear udder height	RUH			5.77	1.38
Temperament	TEM			5.70	1.29
Ease of milking	EOM			5.59	1.11
Dairy character	DAIRY			80.32	6.59
Body	BODY			79.81	5.71
Legs and feet	LAF			77.70	6.86
Mammary	MA			78.24	6.78

In the establishment of maternal families all cows were traced back to founder ancestors using the pedigree registration records database of HUKI. This database holds details of 5 million pedigree registrations from 1960 to the present. All cows used in the analysis were full pedigree animals and were required to belong to maternal families with at least 2 members in the data. The tracing resulted in the 33,325 cows with records being assigned to 10,332 cow families, with 66% in families of less than 5 cows. One cow family was found to have 100 members. Details of the family structure can be seen in Table 4.2.

Table 4.2. *Distribution of cows with records over maternal families*

Family Size	Number of animals with records	Number of families	Average family size	Average number of generations to origin
> 1 and < 5	22031	8881	2.5	5
> 4 and < 10	7299	1202	6.0	6
> 9 and < 20	2566	205	12.5	7
> 19 and < 30	638	26	24.5	7
> 29 and < 40	265	8	33.0	7
> 39 and < 50	260	6	43.0	7
> 49 and < 60	166	3	55.0	8
> 59 and < 70	0	0	-	-
> 69 and < 80	0	0	-	-
> 79	100	1	100.0	8

4.2.2 Data analysis

The analysis was performed using REML VCE (Groeneveld, 1996), fitting a univariate animal model. The pedigree used in the model comprised sire, dam, maternal and paternal grandsire and granddam to give a total of 93728 animals. All animals with records were contemporaries; i.e. no daughter-dam pairs were present in the data. This analysis provides an important enhancement to previous designs that have been used to estimate maternal family variance effects. As previously discussed by Roughsedge *et al.* (1998), using records collected over a number of years increases the heterogeneity of variance of the traits. In addition a design that has daughters and dams with records in the analysis may provide additional phenotypic covariances that the model assigns to maternal lineage but which are in fact not attributable to maternal family. This analysis allowed no such daughter-dam covariance.

4.2.3 Model

$$Y_{ijklmn} = HYV_i + M_j + S_k + SYOB_l + b_1(age) + b_2(age^2) + b_3(HOLP) + a_m + f_n + e_{ijklmn}$$

Where Y_{ijklmn} = trait; HYV_i = fixed effect of herd*year*visit of classifier, (i=1,3495); M_j = month of calving, (j=1,12); S_k = stage of lactation at classification, (k=1,13); $SYOB_l$ = sire year of birth, (l=1,27); age = age at classification; $b_1(age)$ = linear regression of Y on age; $b_2(age^2)$ = quadratic regression of Y on age; HOLP = percentage Holstein; $b_3(HOLP)$ = linear regression of Y on HOLP; a_m = additive genetic effect of the animal, (m=1,93728); f_n = random effect of maternal lineage, (n=1,10332); e_{ijklmn} = residual error. The model was applied to the data with and without the random effect of maternal lineage (ML).

4.2.4 Significance testing

To determine the number of independent traits that were responsible for the total phenotypic variance in type, a principal component analysis was carried out. The input to this was the phenotypic correlation matrix of type traits used in the UK national evaluations. The phenotypic correlations were used because the maternal lineage component that we are estimating is a component of the overall phenotypic variance rather than a component of the additive nuclear variance. Procedures were adopted to determine the number of principal components to retain. Previously the 'latent roots less than one' (Kaiser, 1958), where the latent root is the eigenvalue, has been used as a rejection criteria, (e.g. Vukasinovic *et al.*, (1997)). The rationale of this rejection procedure is that only principal components accounting for more variation than any individual trait are retained. This component rejection criterion was compared with the scree test proposed by Cattell (1966). The basis of the scree test was formalised by Bentler and Yuan (1996). The scree test is used as a means of separating the linearly decreasing small roots from the important large roots. The value before the start of this linear trend is the last component to be retained by the test.

The test statistic used in the analysis was the LRT (Log Likelihood Ratio Test). Under the null hypothesis of no variation due to maternal lineage, the asymptotic distribution of the LRT is $1/2\chi^2(0) + 1/2\chi^2(1)$, (e.g. Stram and Lee, 1994). This implies that for a single LRT, the appropriate P-value for the test statistic is half of the P-value from a $\chi^2(1)$ distribution, or, equivalently, that the threshold for a given type-I error of α is the threshold from a $\chi^2(1)$ pertaining to 2α. For example, for a single trait univariate estimation the threshold value for an experiment wise error rate of $\alpha = 0.05$, is 2.7. However 23 traits are being tested and the number of independent traits being tested can be determined by the use of principal component analysis. The Dunn-Šidák method determines the probability of a type 1 (α) error given k independent traits are being tested, $(1 - \alpha')^k$. Under a LRT, which is distributed $1/2\chi^2(0) + 1/2\chi^2(1)$ the α' must be further divided by 2 giving us $(1 - \alpha'/2)^k$.

The experiment wise error rate is then, $\alpha = 1 - (1 - \alpha'/2)^k$. This can then be solved for $\alpha = 0.05$:

$$\alpha' = 2(1 - (1 - \alpha)^{1/k})$$

If $k = 23$ and $\alpha = 0.05$ this gives us $\alpha' = 0.004$ requiring a LRT test statistic of 8.07 for a significant additional component of variance. Note that for $k = 1$, $\alpha' = 2\alpha$, as expected.

4.3 Results

A summary of the traits used is given in Table 4.1. Brotherstone *et al.*, (1990) found significant between classifier differences in type trait scoring. All type traits recorded by HUKI are therefore adjusted to account for differences in the range of scoring by scaling data from each classifier by the inverse of the corresponding observed phenotypic standard deviation.

Cow families were found to be distributed across a large number of herds within family. For example the largest family numbered 100 cows with records and was distributed across 64 herds. Across all maternal families the number of herds per family was found to be about half the family size.

Overall means and standard deviations of the traits can be seen in Table 4.1. The trait means compare closely to those estimated by Brotherstone *et al.* (1990) except for STA which was seen to increase from 4.45 to 6.45 due to the increased influence of the North American Holstein. Estimates of heritability and the maternal lineage variance components expressed as a proportion of the overall phenotypic variance are given in Table 4.3. The heritability estimates obtained agree closely with the values for linear type trait heritability used in the UK national evaluations (S. Brotherstone). When these

heritability estimates were compared to earlier UK estimates (Brotherstone *et al.*, 1990) some traits showed an increase in heritability, most notably ANG rose from 0.22 to 0.34, RA rose from 0.26 to 0.33 and TSC rose from 0.29 to 0.33.

The estimates obtained for maternal family lineage variance were very low with only STA and BODY having a component above 1% of phenotypic variance attributable to maternal family. The variance component found for BODY was significant at the 5% level given that the 23 type traits were considered independent in the analysis. However we know that the traits are not independent and therefore the significance level estimate is very conservative. In order to determine the number of independent factors accounting for the variance in type, a principal component analysis was performed.

The results of the principal component analysis using the phenotypic correlation matrix can be seen in Table 4.4. The decrease in the eigenvalues from the first component to the last can be shown graphically (Figure 4.1) and from this the number of components that should be retained were determined using a scree analysis (Cattell, 1966). In Figure 4.1 the line fitted to the last 15 eigenvalues accounts for over 97% of their variation, and if the last value is dropped over 99% of the variation is accounted for. A second method of determining the number of components to retain, the 'latent root greater than one' criterion proposed by Kaiser (1958), was also employed. These criteria compare favourably in retaining respectively 7 and 8 components. If we take the more conservative estimate of 8 principal components being responsible for the overall variation in type traits then using the Dunn-Šidák method the level of significance of the stature and body traits are increased to probability levels of 0.059 and 0.016 respectively.

Table 4.3. *Heritability and maternal family variance component (±s.e.) estimates*

TRAIT	Without maternal		With Maternal Lineage				
	h^2		h^2		f^2		LRT
TSC	0.333	(0.019)	0.333	(0.019)	0.000	(0.000)	0.00
STA	0.516	(0.021)	0.499	(0.020)	0.011	(0.003)	5.93
CW	0.229	(0.016)	0.229	(0.015)	0.000	(0.000)	0.00
BD	0.358	(0.020)	0.352	(0.018)	0.006	(0.003)	2.10
ANG	0.337	(0.018)	0.331	(0.018)	0.009	(0.003)	1.74
RA	0.331	(0.018)	0.330	(0.021)	0.002	(0.005)	0.13
RW	0.292	(0.017)	0.292	(0.017)	0.000	(0.000)	0.00
RLS	0.159	(0.013)	0.156	(0.012)	0.005	(0.003)	1.66
FA	0.139	(0.012)	0.138	(0.013)	0.001	(0.003)	0.05
FUA	0.188	(0.015)	0.185	(0.013)	0.004	(0.003)	1.22
US	0.183	(0.015)	0.177	(0.014)	0.009	(0.003)	4.56
UD	0.318	(0.017)	0.318	(0.016)	0.000	(0.000)	0.00
TPR	0.302	(0.018)	0.297	(0.017)	0.005	(0.002)	1.60
TPS	0.290	(0.017)	0.289	(0.014)	0.001	(0.003)	0.06
TL	0.322	(0.018)	0.322	(0.017)	0.000	(0.000)	0.00
CS	0.235	(0.015)	0.235	(0.016)	0.000	(0.000)	0.00
RUH	0.334	(0.015)	0.334	(0.015)	0.000	(0.000)	0.00
TEM	0.104	(0.011)	0.104	(0.011)	0.000	(0.000)	0.00
EOM	0.098	(0.010)	0.096	(0.010)	0.006	(0.003)	1.46
DAIRY	0.376	(0.020)	0.376	(0.018)	0.000	(0.000)	0.00
BODY	0.381	(0.021)	0.361	(0.015)	0.015	(0.004)	8.29
LAF	0.156	(0.013)	0.156	(0.012)	0.000	(0.000)	0.00
MA	0.226	(0.016)	0.226	(0.016)	0.000	(0.000)	0.00

Table 4.4. *Eigenvalues of principal components of type trait phenotypic correlation matrix*

Component	Eigenvalue	As Percentage[a]	Cum. Percentage
1	5.03	21.88	21.88
2	2.71	11.80	33.69
3	2.24	9.74	43.43
4	1.62	7.05	50.48
5	1.25	5.42	55.90
6	1.21	5.25	61.15
7	1.06	4.62	65.77
8	0.97	4.21	69.98
9	0.77	3.35	73.33
10	0.75	3.28	76.61
11	0.71	3.08	79.70
12	0.68	2.95	82.65
13	0.61	2.65	85.30
14	0.58	2.52	87.81
15	0.55	2.39	90.20
16	0.48	2.11	92.31
17	0.40	1.73	94.04
18	0.36	1.57	95.60
19	0.31	1.35	96.96
20	0.28	1.23	98.18
21	0.25	1.09	99.27
22	0.16	0.69	99.95
23	0.01	0.05	100.00

[a]Percentage of total aggregate variance of 23 type traits explained by principal component

Table 4.5. *Eigenvectors of first eight principal components*

	First 8 principal component eigenvectors							
Traits	1	2	3	4	5	6	7	8
TSC	-0.42	0.03	-0.06	-0.06	-0.04	0.07	-0.04	-0.02
STA	-0.25	-0.30	0.10	0.05	0.32	-0.15	-0.15	0.04
CW	0.00	-0.49	-0.05	0.12	-0.17	0.11	0.01	-0.04
BD	-0.22	-0.38	0.19	0.08	0.00	0.03	-0.04	0.04
ANG	-0.30	0.21	0.31	-0.08	0.12	-0.04	-0.03	0.07
RA	0.05	-0.04	0.04	-0.07	0.09	-0.22	-0.52	-0.77
RW	-0.12	-0.37	0.05	0.13	-0.02	0.03	0.03	0.13
RLS	0.02	0.10	0.25	0.52	-0.02	0.00	0.00	0.02
FA	-0.06	-0.14	-0.22	-0.53	0.02	0.04	-0.01	0.08
FUA	-0.15	0.02	-0.42	0.20	0.19	-0.06	0.09	0.02
US	-0.24	0.14	-0.13	0.11	-0.09	0.15	0.08	-0.22
UD	-0.10	0.06	-0.38	0.15	0.44	-0.23	-0.07	0.09
TPR	-0.17	0.07	-0.20	0.18	-0.34	0.14	-0.32	-0.03
TPS	-0.16	-0.01	0.15	-0.03	-0.35	0.32	0.06	-0.14
TL	-0.01	-0.12	0.12	-0.07	0.24	0.06	0.66	-0.49
CS	0.17	-0.37	-0.26	0.08	-0.19	0.01	0.03	-0.08
RUH	-0.26	0.08	-0.22	0.11	0.07	0.12	0.23	-0.13
DAIRY	-0.33	0.16	0.31	-0.09	0.02	-0.03	-0.02	0.03
BODY	-0.30	-0.28	0.16	0.02	0.13	-0.09	-0.11	0.08
LAF	-0.18	-0.04	-0.13	-0.48	-0.07	0.04	-0.04	0.05
MA	-0.34	0.12	-0.26	0.14	-0.14	0.10	0.02	-0.09
TEM	-0.10	-0.02	-0.01	0.00	-0.34	-0.56	0.24	-0.08
EOM	-0.10	0.03	-0.02	-0.01	-0.35	-0.59	0.12	0.07

Values highlighted are those > | 0.30 |

4.4 Discussion

The analysis of type traits of UK Holstein Friesian dairy cows demonstrated a significant component of phenotypic variance attributable to maternal lineage for the body and stature traits. Phenotypic and genetic correlations of 0.93 and 0.64 between body and stature were found for this data and it is not surprising that if one trait has a significant maternal lineage component then so does the other. However, the magnitudes of the variance components estimated were not large and the majority of the type traits were not found to have significant variance components attributable to maternal lineage. This provides evidence counter to the belief of some breeders that certain cow families are of superior type to others, irrespective of the bulls used.

When production traits are being considered it is easy to see how hypothesising that a mechanism encoding proteins involved in the metabolism of energy could be responsible for a component of phenotypic variance. However, in considering type traits it is not obvious what we would expect to be significantly affected by mitochondrial DNA. If we consider the trait most directly connected with the energetic process, condition score (CS), then it is possible to see a possible mitochondrial influence on the trait. However, CS was not found to have a component of variance attributable to maternal family. Certainly all type traits are the result of physiological processes and most show a reasonable level of heritability. In this study little or no influence of maternal family in addition to the nuclear genetic effects was found.

One concern with a study looking for maternal lineage effects in a commercial pedigree population is the bias that could be present due to preferential treatment (PT). If a producer is aware that a cow is from an elite lineage they are liable to accord that animal extra management. This extra management may result in increased yield and better condition. In terms of type traits however it is not clear what the effect of preferential

management practises would be. Several studies have looked at the effect that PT of cows has on the estimation of sire PTA's for production traits, both from the perspective of bull-dam selection and bull daughter treatment. Most of these studies simulate treatment effects (e.g. Kuhn *et al.*, 1995; Weigel *et al.,* 1994). However some studies have attempted to estimate PT using production data (e.g. Graham *et al.*, 1991). These studies have largely concluded that only a high level of preferential treatment would lead to a bias in sire rankings. The way in which bias is applied is also of significance. Weigel *et al.* (1994) used several strategies of both the applications of bias and in the selection of bull dams in a simulated dairy population over 13 years of selection. They found that bias attributable to maternal family membership resulted in only 2% of cows being biased. However, this type of bias was consistently found to be the most damaging to bull dam selection. Bias applied to cow family was also difficult to detect using within herd phenotypic variance as a measure of preferential treatment. Within herd phenotypic yield variation is the standard test for detecting preferential treatment in production data (e.g. Graham *et al.*, 1991). In this study it was found that cow families were distributed over a number of herds. If cow families were thought to be of superior type and were afforded differential levels of management in some herds, then an upward bias in the within family variation would be one possible outcome. To test the hypothesis that PT was being given to cows, on the basis of cow family, a nested REML model was run for a sample of type traits using the same model as described previously but with a herd by family random effect included. A zero component of variance attributable to this random herd by family effect was consistently found and the hypothesis that preferential treatment causes a bias in the estimation of maternal lineage variance component estimation was rejected.

Twenty three traits were tested in this study. If we make the assumption that these traits are all independent then the threshold required for a 5% significance level, using the Dunn-Šidák method, increases to 8.07. This is no doubt being too conservative given that these are not all independent traits. Essentially linear type traits can be broadly categorised into three areas, body, legs and feet and mammary traits. In addition to this we have four

composite traits, body, dairy, mammary and legs and feet, a total score trait and measures of temperament and ease of milking. The composites are highly correlated with the linear categories. The information provided by the principal component analysis was used to determine the number of independent traits that were responsible for the variation in type. This number of independent traits is required to establish the level of significance of the LRT test statistics obtained in the REML analysis. The use of the scree test to ascertain the number of independent components that adequately describe the variance in type is visually appealing. Bentler and Yuan (1996) developed a formal test statistic for the scree plot. They contend that many researchers discard eigenvalues on the rational of the rejected eigenvalues being of equal and small magnitude. This has not been found to be the case as the smaller eigenvalues are found to exhibit steady linear decrease in most data sets. A test was developed based on this linearity. The principal components corresponding to the small and linearly decreasing eigenvalues may just reflect rounding errors or other instrumental errors. Vukasinovic *et al.* (1997) used principal component analysis to estimate the number of type trait principal components accounted for most of the variance in type. 18 linear type traits of Brown Swiss cows were used. However, the traits scored were not identical to the current study. It was found that the first 5 components accounted for 58% of the phenotypic variance compared to 56% in the current study. When the genetic correlation matrix was used to determine principal components the first 5 components accounted for 72% of the variance compared to 74% in the study of Vukasinovic *et al.* (1997). If the eigenvectors of the initial 8 principal components are examined we can establish the traits that are contributing to these components (Table 4.5). For example, the first component is composed of total score and a combination of the composite scores, the second of body traits, the third of udders and dairy character and the fourth of leg and feet traits. In the Brown Swiss

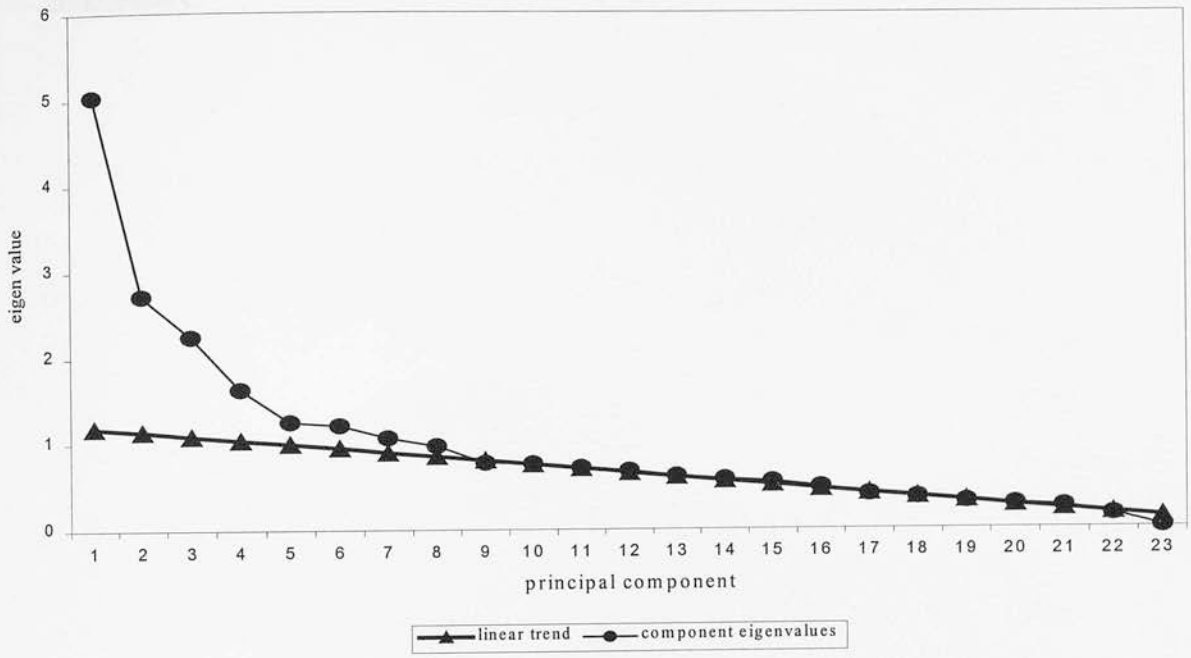

Figure 4.1 Scree plot of eigenvalues of 23 principal components showing linear relationship of rejected components.

study the component eigenvalues < 1 rejection criteria was used, and resulted in 5 components being retained. The rationale of this test is that eigenvalues of less than one account for less variation than that explained by a single trait. In the current study applying the above rejection criteria we would retain only seven components, one less than retained by the scree test.

Evidence for a component of maternal lineage variance in type traits of Holstein Friesian dairy cattle was found in only two traits, body and stature. It would appear that giving greater consideration to cows coming from cow families of 'good type' in selection

decisions in addition to individual animal estimated breeding values for type is unnecessary.

Chapter 5

Effects of cow families on production traits in dairy cattle

5.1 Introduction

In mammals the mitochondria are almost exclusively inherited from the maternal parent (Hutchinson *et al.* 1974) and hence, with the exception of mutational events, all animals of a maternal lineage have identical mtDNA. It is therefore theoretically possible to estimate a component of variance due to maternal lineage, which in turn is assigned to mtDNA. A theoretical experimental design to separate cytoplasmic effects and all interactions was proposed by Beavis *et al.* (1987). In genetic analysis of dairy cattle higher estimates of heritability have been obtained from daughter-dam regression than from paternal half-sib analysis (e.g. Seykora and McDaniel, 1983; Visscher and Thompson, 1992). This suggests that there may be a mechanism of inheritance, in addition to nuclear genetic inheritance, which is not being accounted for in current evaluations. The method of Beavis *et al.* (1987) is not feasible for dairy cattle but it has been shown (Southwood *et al.*, 1989) that with field data the use of an animal model to account for all nuclear contributions will enable the separation of cytoplasmic inheritance if it is present. Southwood *et al.* (1989) also demonstrated that if maternal genetic effects were simulated, and the data analysed using an incorrect model that accounted only for additive and cytoplasmic effects, but not maternal genetic effects, then a small cytoplasmic variance component would be detected. Most of the maternal genetic variance would be partitioned into additive genetic variance.

Several analyses have estimated the level of phenotypic variance attributable to maternal lineage using an animal model (e.g. Boettcher and Gibson, 1997; Schnitzenlehner and Essl, 1999, Chapter 2). These analyses used the animal model and fitted maternal lineage as a random effect. None of the above analyses demonstrated a component of variance for production traits in dairy cattle attributable to maternal lineage variance. Persistency and a herd life trait were shown however to have a significant component of maternal lineage

variance (Schnitzenlehner and Essl, 1999). A significant 1.5% component of maternal lineage variance was also estimated for the composite dairy type trait, body (Chapter 4).

The aim of this study was to estimate the components of phenotypic variance of production traits of the UK Holstein Friesian dairy population that are attributable to maternal lineage. Using test day records it is possible to explore aspects of the lactation curve in terms of persistency and different periods of production. Given the hypothesis that mitochondrial DNA is connected with energetic processes it is not unreasonable to speculate that maternal lineage variance is likely to be highest at the time of peak production, i.e. early lactation, and also in the maintenance of production over a lactation, i.e. persistency. The current study also aims to overcome the problem highlighted by the study of Southwood *et al.*, (1989) of attributing direct maternal effects to cytoplasmic inheritance. This is achieved by using a contemporary data set that includes no daughter-dam pairs with records. The study also investigates the preferential treatment of cow families.

5.2 Material and Methods

A total of 55,230 first lactation Holstein Friesian pedigree cows were available for the analysis. Type records from these cows were used previously to investigate maternal lineage variance (Chapter 4). Two data restrictions based on family size were applied. The first restricted family size to greater than one and the second to greater than four cows per maternal lineage, which resulted in data sets of 36,320 and 10,598 records available for analysis respectively. All cows calved between 1996 and 1998.

Information on 305 day yields and test day records were available. A summary of the traits analysed is in Table 5.1.

Table 5.1. *Trait information for full data set and for restricted data sets.*

Trait	Abbreviation	Mean	SD	Number Records
Milk (kg)	MLK	6577.39	1373.92	55230
Fat (kg)	FAT	267.03	52.87	55230
Protein (kg)	PRT	215.51	43.66	55230
Fat (%)	FATP	4.09	0.43	55230
Protein (%)	PRTP	3.29	0.19	55230
Milk (kg)	MLK[a]	6622.18	1387.29	36320
Fat (kg)	FAT[a]	268.54	53.19	36320
Protein (kg)	PRT[a]	216.89	44.05	36320
Fat (%)	FATP[a]	4.09	0.43	36320
Protein (%)	PRTP[a]	3.28	0.19	36320
Milk (kg)	MLK[b]	6765.83	1430.66	10596
Fat (kg)	FAT[b]	273.31	54.06	10596
Protein (kg)	PRT[b]	221.26	45.25	10596
Fat (%)	FATP[b]	4.07	0.43	10596
Protein (%)	PRTP[b]	3.28	0.19	10596
Avg. Test 1-3 Milk (kg)	T1-3 MLK[a]	24.83	5.38	93486
Avg. Test 1-3 Fat (kg)	T1-3 FAT[a]	0.98	0.23	93486
Avg. Test 1-3 Protein (kg)	T1-3 PRT[a]	0.78	0.17	93486
Avg. Test 1-3 Fat (%)	T1-3 FATP[a]	3.97	0.62	93486
Avg. Test 1-3 Protein (%)	T1-3 PRTP[a]	3.17	0.27	93486
Avg. Test 1-3 Milk (kg)	T1-3 MLK[b]	25.27	5.50	28185
Avg. Test 1-3 Fat (kg)	T1-3 FAT[b]	0.99	0.23	28185
Avg. Test 1-3 Protein (kg)	T1-3 PRT[b]	0.80	0.17	28185
Avg. Test 1-3 Fat (%)	T1-3 FATP[b]	3.96	0.62	28185
Avg. Test 1-3 Protein (%)	T1-3 PRTP[b]	3.16	0.27	28185
Persistency[c]	PERSIS3[a]	77.45	17.60	26572
SD of Milk Tests[c]	SD3[a]	3.97	1.48	26572
Persistency[c]	PERSIS3[b]	77.77	16.66	7241
SD of Milk Tests[c]	SD3[b]	3.87	1.51	7241

[a] Data set restricted to > 1 cows per maternal family

[b] Data set restricted to > 4 cows per maternal family

[c] Measured over first 300 days of lactation

5.2.1 Traits

The traits were 305 day milk (MLK), fat (FAT) and protein (PRT) yield, fat (FATP) and protein (PRTP) percent. Also analysed were test days one to three for milk (T1-3MLK), fat (T1-3FAT), protein (T1-3PRT) yield and fat (T1-3FATP) and protein (T1-3PRTP) percent.

Two derived traits of persistency measures were also analysed. The first was the ratio of the first 100 days of production to the last 100 day of production in the first 300 days of milk production (P3) (Sölkner and Fuchs, 1987). The second was standard deviation of test day milk yield over the first ten test days (SD3) suggested by Sölkner and Fuchs (1987).

$$P3 = \frac{Milk\ yield\ third\ 100\ days}{Milk\ yield\ first\ 100\ days} \times 100$$

SD3 = Standard deviation of test-day milk yields (300 days)

In the calculation of P3 the exponential curve of Wilmink (1987) was used to calculate daily yield.

$$Y_t = a + be^{-kt} + ct \qquad (5.1)$$

where Y_t is the yield on day t of lactation. Curve parameters a, b and c were estimated for each cow and parameter k was fixed at 0.068 (following White *et al*, 1999). Fitting the curve resulted in the rejection of about 5% of records due to a poor fit. The Wilmink curve is a four parameter curve but the fourth parameter is generally set to a constant for convenience. It was found to perform better than other three parameter models by Olori *et al.* (1999). The measure P3 was calculated (Eq. 5.2) using the fitted values ($\hat{Y}_t$) from Eq. (5.1)

$$P3 = (\sum_{t=201}^{t=300} \hat{Y}_t / \sum_{t=1}^{t=100} \hat{Y}_t) \times 100 \qquad (5.2)$$

When test day records were used in both the derivation of persistency traits and in their direct use as traits of interest a restriction to the data set was applied to restrict the third test day to the third month of lactation. This was to ensure that animals with missing test days in the first three test days were excluded from the analysis.

5.2.2 Maternal Families

Maternal families were established by tracing maternal pedigrees of all cows with records on the Holstein UK and Ireland (HUKI) database. This database holds 5 million pedigree records, back to 1960. All cows used in the analysis were full pedigree registered cows. No cows used were recent grade-up cows, i.e. all were at least four generations from a non-pedigree dam. This was done in order to avoid small maternal families. However of the 55,230 cows with records, about 40% belonged to a single cow family, i.e. one cow with a record tracing back to one maternal ancestor. The maternal family structure is shown in Table 5.2. Previous studies looking for maternal lineage variance components using field data have made data restriction based on size of cow family, e.g. Boettcher and Gibson (1997) and Schnitzenlehner and Essl (1999) made restrictions of greater than or equal to two and ten cows per cow family respectively. In the data set used for the current analysis contemporary data was used which limits the number of large cow families but provides benefits in the structure of the data for the detection of true maternal lineage variance effects. With contemporary data most relationships are no closer than cousins sharing 1/16 of nuclear genetic material but maternal cousins share 100% of their mtDNA, enabling a better opportunity to separate variance attributable to mtDNA.

Table 5.2. *Distribution of cows with records over maternal families*

Family Size	Number animals with records	Number of families	Average family size	Average number generations to origin
> 1 and < 5	25716	10292	2.5	5
> 4 and < 10	7931	1314	6.0	6
> 9 and < 15	1331	118	11.5	6
> 14 and < 20	557	34	16.5	6
> 19 and < 30	500	21	24.0	7
> 29 and < 40	139	4	35.0	7
> 39 and < 50	89	2	44.5	7
> 49	57	1	57.0	8

5.2.3 Data analysis

The analysis was performed using REML VCE (Groeneveld, 1996) fitting a univariate animal model. The pedigree used in the model with greater than one record per maternal lineage comprised sire, dam, maternal and paternal grandsire and granddam to give a total of 100,643 animals. The animals with records were all contemporaries; i.e. there were no daughter-dam or daughter-granddam pairs in the data. This contemporary design was used in Chapter 4. The use of such a design is intended specifically to remove additional covariance from the data structure that can be attributed to daughter-dam relationships but more generally to reduce the degree of additive relationship between individuals within cow families. By reducing additive genetic relationship the within family similarity due to mtDNA becomes more apparent. In addition to reducing additive genetic relationship and removing the extra covariance structure between daughters and dams, the use of non-contemporary data reduces the heterogeneity of variance of the traits (Roughsedge *et al.*, 1998).

Three models were fitted for the 2 or more records per maternal lineage data sets. Model 1 was fitted for 305 day milk, fat and protein yield traits and fat and protein percentage. Model 2 was fitted for the test day yield traits, with a random effect for permanent environment where the first three test days were fitted. Model 3 was fitted to the two persistency traits, SD3 and P3. When data was restricted to five or more records per maternal lineage then model 1 was altered to fit both herd by year as a fixed effect and month of calving as a separate fixed effect due to small effects groups.

5.2.4 Model

Model 1

$$Y_{ijkl} = HYS_i + b_1(age) + b_2(age^2) + b_3(HOLP) + a_j + f_k + hf_l + e_{ijkl}$$

Model 2

$$Y_{ijklmn} = H_i + TD_j + b_1(age) + b_2(age^2) + b_3(dim) + b_4(dim^2) + b_5(HOLP) + a_k + pe_l + f_m + hf_n + e_{ijklmn}$$

Model 3

$$Y_{ijkl} = HYS_i + b_1(age) + b_2(age^2) + b_3(HOLP) + b_4(milk) + a_j + f_k + hf_l + e_{ijkl}$$

Where Y_{ijklmn} = trait; HYS_i = fixed effect of herd*year*season of first test day, (i = 7170); H_i = fixed effect of herd, (i = 2774), season was divided into three four month blocks starting with January; TD_j = fixed effect of year*month of test day, (j = 1,29); age = age at classification; $b_1(age)$ = linear regression of Y on age; $b_2(age^2)$ = quadratic regression of Y on age; HOLP = percentage Holstein; $b_3(HOLP)$ = linear regression of Y on HOLP; dim = number days in milk; $b_3(dim)$ = linear regression of Y on dim; $b_4(dim^2)$ = quadratic

regression on dim; milk = 305 day milk yield; b_4(milk) = linear regression of Y on milk; a_m = additive genetic effect of the animal, (m = 1,100640); pe_l = permanent environmental effect of animal; f_n = random effect of maternal lineage, (n = 1,11789); hf_l = random effect of herd*family, (l = 1,17475); e_{ijklmn} = residual error. The model was applied to the data with and without the random effect of maternal lineage (f). The model was also applied with a herd by family random effect (hf). Herd is a fixed effect in the model so the hf effect accounts for both a family effect and a herd by family interaction effect. In order to test for a significant interaction effect the model with herd by maternal family was tested against the model with maternal family.

It is not straightforward to establish the threshold value required for an extra component of variance to be significant in an analysis where several traits are being tested and several components of variance are being fitted. It would appear wise to err on the side of caution in such a situation. It is known when dealing with yield traits that there is a strong correlation structure between them. Persistency traits were estimated to be highly correlated with milk yield by Solkner and Fuchs (1987) but it was shown that the sign of the correlation depended on the chosen definition of persistency, i.e. SD3 is negatively correlated with milk production and P3 is positively correlated. In order to accommodate this positive or negative correlation persistency is regressed on milk yield in the model and is thus independent of yield.

5.2.4 Test Statistic

The test statistic used in the analysis was the LRT (Log Likelihood Ratio Test). Under the null hypothesis of no variation due to maternal lineage, the asymptotic distribution of the LRT is $1/2\chi^2(0) + 1/2\chi^2(1)$, (e.g. Stram and Lee, 1994). This implies that for a single LRT, the appropriate P-value for the test statistic is half of the P-value from a $\chi^2(1)$ distribution, or, equivalently, that the threshold for a given type-I error of α is the threshold from a $\chi^2(1)$ pertaining to 2α. For example, for a single trait univariate estimation the threshold value for an experiment wise error rate of $\alpha = 0.05$, is 2.7. However more than one

independent test is being carried out. The Dunn-Šidák method determines the probability of a type 1 (α) error given k independent traits are being tested, $(1 - \alpha')^k$. Under a LRT, which is distributed $1/2\chi^2(0) + 1/2\chi^2(1)$ the α' must be further divided by 2 giving us $(1 - \alpha'/2)^k$. The experiment wise error rate is then, $\alpha = 1 - (1 - \alpha'/2)^k$. This can then be solved for $\alpha = 0.05$:

$$\alpha' = 2(1 - (1 - \alpha)^{1/k})$$

The most extreme case would be to consider that all traits investigated were independent and that the two extra variance components being investigated, i.e. family and the family by herd interaction were independent tests. In this scenario five 305 day yield traits, two persistency traits and five test day traits are being tested giving 12 traits and 24 independent tests. However, given that there is a very strong correlation structure between the traits being investigated fewer independent tests are being used. It is not possible to isolate the exact level of the threshold test statistic but to set it too high would seriously reduce the power of detecting a variance component. If $k = 12$ and $\alpha = 0.05$ this gives $\alpha' = 0.008$ requiring a LRT test statistic of 7.03 for a significant additional component of variance. Note that for $k = 1$, $\alpha' = 2\alpha$, as expected.

5.3 Results

A summary of the traits analysed can be seen in Table 5.1. The traits are given with their mean and standard deviation. The numbers of records, resulting from the various data editing procedures described, are also shown.

Results of the analyses can be seen in Tables 5.3 to 5.6. Tables 5.3 and 5.4 show the results of fitting the random maternal lineage effect. The full, unedited data set was used for 305 day traits in order to determine that the estimation of heritability was not being affected by the data editing procedure. It can be seen that the estimation of heritability was

not significantly affected by the maternal family data restrictions imposed. Using the full data set no significant component of phenotypic variance attributable to maternal lineage was found (Table 5.3).

Table 5.3. *Heritability (±SE) estimates etc. for model 1 with and without maternal lineage fitted.*

	Without maternal lineage	With Maternal Lineage		
Trait	h^2	h^2	f^2	LRT
MLK	0.436 (0.017)	0.436 (0.015)	0.001 (0.001)	0.04
FAT	0.403 (0.017)	0.399 (0.018)	0.004 (0.003)	0.69
PRT	0.420 (0.017)	0.415 (0.018)	0.002 (0.004)	0.17
FATP	0.644 (0.019)	0.638 (0.020)	0.006 (0.003)	1.49
PRTP	0.655 (0.020)	0.652 (0.018)	0.002 (0.003)	0.10
MLK[a]	0.443 (0.020)	0.440 (0.019)	0.002 (0.004)	0.20
FAT[a]	0.382 (0.019)	0.374 (0.018)	0.007 (0.004)	2.60
PRT[a]	0.415 (0.019)	0.411 (0.019)	0.003 (0.003)	0.42
FATP[a]	0.655 (0.021)	0.645 (0.021)	0.006 (0.004)	1.95
PRTP[a]	0.666 (0.021)	0.660 (0.021)	0.004 (0.004)	0.62
PERSIS3[a]	0.119 (0.013)	0.117 (0.013)	0.006 (0.007)	0.49
SD3[a]	0.160 (0.015)	0.159 (0.016)	0.002 (0.005)	0.09
MLK[b]	0.403 (0.032)	0.402 (0.032)	0.002 (0.005)	0.04
FAT[b]	0.386 (0.032)	0.374 (0.030)	0.012 (0.006)	2.50
PRT[b]	0.387 (0.033)	0.386 (0.028)	0.001 (0.006)	0.02
FATP[b]	0.636 (0.035)	0.627 (0.036)	0.008 (0.006)	1.00
PRTP[b]	0.608 (0.033)	0.608 (0.033)	0.000 (0.000)	0.00
PERSIS3[b]	0.136 (0.030)	0.104 (0.028)	0.044 (0.011)	11.21
SD3[b]	0.115 (0.028)	0.115 (0.028)	0.000 (0.000)	0.00

[a] Data set restricted to > 1 cows per maternal family

[b] Data set restricted to > 4 cows per maternal family

When the data was restricted to families of two or more cows the log likelihood ratio test statistic was seen to rise for fat yield. However, the level of test statistic was not significant at a 5% type 1 error rate using the one independent test threshold value for the LRT of 2.7, and the component was less than 1%. The results from the analysis of 305 day yield data are comparable to those obtained by Boettcher and Gibson (1997) using field data from the Canadian Holstein herd. When the further data restriction of five or more records per cow family was applied to 305 day yield traits no change was seen.

The persistency traits present a different picture (Table 5.3). When the data editing allowed records in families of two or more cows in the analysis then there was no significant component of variance attributable to maternal lineage. However, when further restrictions were placed on maternal family size the P3 trait was estimated to have a 4.4% variance component attributable to maternal lineage and this component was found to be highly significant. Indeed the LRT value would provide a 5% experiment wise error rate given 125 independent tests. If we assume that we are undertaking at most 12 independent test which is still conservative then the probability of a type 1 error for the P3 trait is 0.0048. However, no maternal lineage component was estimated for the SD3 trait.

Yields from tests one to three, fitted as a repeatability model (Table 5.4), did not show any significant maternal lineage variance component. These yield records were also individually investigated using a univariate model and again no significant maternal lineage component was estimated.

The results of fitting the herd by maternal family interaction are shown in Tables 5.5 and 5.6. The model to fit the interaction term fitted a herd*maternal family random effect in addition to the animal random effect. This random effect accounted for both a family effect and an interaction effect. In order to test for the interaction component this model was tested against the model that fitted only animal and maternal family. For the 305 day yield traits and the persistency traits, whichever data edit was applied, no significant interaction component was detected. However the test days one to three repeatability

model (Table 5.6) detected a 1.1% and 0.7% component of phenotypic variance attributable to herd*maternal lineage for milk and protein yield respectively, when greater than one cow per maternal lineage was allowed. The further data restriction of greater than or equal to five cows per maternal lineage increased the component detected to 1.5% and 1.7% for milk and protein yield respectively. The individual test days were investigated as univariate traits and test day two was found to have consistent estimates of zero for the interaction component. However test day one was estimated to have a 2.6% and 3.2% interaction variance component for milk and protein yield respectively with test statistics of 4.64 and 6.92 (Table 5.7). The significance of these components however must be questioned. Suppose that there are only five independent tests being carried out then for a 5% experiment wise error rate (α=0.05) a threshold test statistic of 5.41 would be required. All but one of the above, the test day one protein yield herd*maternal family variance component, are below this level of significance. The test day one protein yield trait interaction component is significant at α=0.05 even if as many as 11 independent tests are being considered.

Table 5.4. *Heritability (±s.e.) estimates etc. for model 2 with and without maternal lineage fitted.*

	Without maternal lineage		With Maternal Lineage			
Trait[c]	h^2	p^2	h^2	p^2	f^2	LRT
T1-3 MLK[a]	0.238 (0.012)	0.371 (0.011)	0.233 (0.012)	0.371 (0.011)	0.005 (0.002)	2.30
T1-3 FAT[a]	0.176 (0.010)	0.270 (0.009)	0.176 (0.010)	0.270 (0.009)	0.000 (0.001)	0.00
T1-3 PRT[a]	0.189 (0.011)	0.326 (0.010)	0.188 (0.011)	0.326 (0.010)	0.001 (0.002)	0.14
T1-3 FATP[a]	0.232 (0.010)	0.078 (0.008)	0.231 (0.010)	0.078 (0.008)	0.000 (0.001)	0.02
T1-3 PRTP[a]	0.246 (0.010)	0.177 (0.008)	0.246 (0.010)	0.177 (0.008)	0.000 (0.000)	0.00
T1-3 MLK[b]	0.226 (0.022)	0.319 (0.020)	0.223 (0.020)	0.319 (0.018)	0.003 (0.004)	0.22
T1-3 FAT[b]	0.163 (0.017)	0.284 (0.015)	0.163 (0.017)	0.284 (0.015)	0.000 (0.000)	0.00
T1-3 PRT[b]	0.207 (0.020)	0.316 (0.018)	0.206 (0.019)	0.316 (0.018)	0.001 (0.004)	0.03
T1-3 FATP[b]	0.203 (0.017)	0.104 (0.015)	0.203 (0.016)	0.104 (0.013)	0.000 (0.002)	0.00
T1-3 PRTP[b]	0.213 (0.018)	0.212 (0.016)	0.212 (0.019)	0.212 (0.017)	0.001 (0.004)	0.07

[a] Data set restricted to > 1 cows per maternal family

[b] Data set restricted to > 4 cows per maternal family

[c] Traits are test day 1 to 3 of lactation

Table 5.5. *Heritability (±s.e) estimates etc. for model 1 with maternal lineage fitted or maternal lineage + herd * maternal lineage fitted.*

	Without maternal lineage		With Maternal Lineage		
Trait	h^2	f^2	h^2	f^2+hf^2	LRT
MLK	0.436 (0.015)	0.001 (0.001)	0.435 (0.015)	0.002 (0.006)	0.03
FAT	0.399 (0.018)	0.004 (0.003)	0.399 (0.017)	0.004 (0.004)	0.00
PRT	0.415 (0.018)	0.002 (0.004)	0.415 (0.018)	0.003 (0.003)	0.08
FATP	0.638 (0.020)	0.006 (0.003)	0.638 (0.019)	0.006 (0.004)	0.00
PRTP	0.652 (0.018)	0.002 (0.003)	0.644 (0.019)	0.009 (0.004)	2.03
MLK[a]	0.440 (0.019)	0.002 (0.004)	0.437 (0.017)	0.005 (0.005)	0.38
FAT[a]	0.374 (0.018)	0.007 (0.004)	0.373 (0.016)	0.008 (0.006)	0.10
PRT[a]	0.411 (0.019)	0.003 (0.003)	0.408 (0.018)	0.006 (0.005)	0.43
FATP[a]	0.645 (0.021)	0.006 (0.004)	0.645 (0.020)	0.006 (0.005)	0.00
PRTP[a]	0.660 (0.021)	0.004 (0.004)	0.651 (0.022)	0.010 (0.004)	2.13
PERSIS3[a]	0.117 (0.013)	0.006 (0.007)	0.117 (0.013)	0.006 (0.008)	0.00
SD3[a]	0.159 (0.016)	0.002 (0.005)	0.157 (0.015)	0.015 (0.008)	2.00
MLK[b]	0.402 (0.032)	0.002 (0.005)	0.394 (0.030)	0.010 (0.008)	0.92
FAT[b]	0.374 (0.030)	0.012 (0.006)	0.369 (0.031)	0.018 (0.008)	0.39
PRT[b]	0.386 (0.028)	0.001 (0.006)	0.376 (0.030)	0.012 (0.008)	1.33
FATP[b]	0.627 (0.036)	0.008 (0.006)	0.672 (0.033)	0.008 (0.006)	0.00
PRTP[b]	0.608 (0.033)	0.000 (0.000)	0.608 (0.036)	0.000 (0.000)	0.00
PERSIS3[b]	0.104 (0.028)	0.044 (0.011)	0.102 (0.026)	0.051 (0.012)	0.48
SD3[b]	0.115 (0.028)	0.000 (0.000)	0.115 (0.028)	0.000 (0.000)	0.00

[a] Data set restricted to > 1 cows per maternal family

[b] Data set restricted to > 4 cows per maternal family

Table 5.6. *Heritability (±SE) estimates etc. for model 2 with maternal lineage or maternal lineage + herd*maternal lineage fitted.*

Trait[c]	Without maternal lineage		With maternal lineage				
	h^2	p^2	f^2	h^2	p^2	$f^2 + hf^2$	LRT
T1-3 MLK[a]	0.233 (0.012)	0.371 (0.011)	0.005 (0.002)	0.228 (0.012)	0.370 (0.003)	0.011 (0.011)	3.56
T1-3 FAT[a]	0.176 (0.010)	0.270 (0.009)	0.000 (0.001)	0.176 (0.009)	0.270 (0.008)	0.000 (0.002)	0.00
T1-3 PRT[a]	0.188 (0.011)	0.326 (0.010)	0.001 (0.002)	0.184 (0.010)	0.325 (0.009)	0.007 (0.003)	2.68
T1-3 FATP[a]	0.231 (0.010)	0.078 (0.008)	0.000 (0.001)	0.321 (0.010)	0.078 (0.008)	0.000 (0.002)	0.00
T1-3 PRTP[a]	0.246 (0.010)	0.177 (0.008)	0.000 (0.000)	0.245 (0.010)	0.177 (0.009)	0.002 (0.002)	0.26
T1-3 MLK[b]	0.223 (0.020)	0.319 (0.018)	0.003 (0.004)	0.213 (0.021)	0.318 (0.018)	0.015 (0.006)	3.53
T1-3 FAT[b]	0.163 (0.017)	0.284 (0.015)	0.000 (0.000)	0.163 (0.015)	0.284 (0.014)	0.000 (0.000)	0.00
T1-3 PRT[b]	0.206 (0.019)	0.316 (0.018)	0.001 (0.004)	0.194 (0.020)	0.313 (0.017)	0.017 (0.006)	4.86
T1-3 FATP[b]	0.203 (0.016)	0.104 (0.013)	0.000 (0.002)	0.203 (0.016)	0.010 (0.014)	0.000 (0.002)	0.00
T1-3 PRTP[b]	0.212 (0.019)	0.212 (0.017)	0.001 (0.004)	0.211 (0.018)	0.212 (0.016)	0.003 (0.004)	0.09

[a] Data set restricted to > 1 cows per maternal family

[b] Data set restricted to > 4 cows per maternal family

[c] Traits are test day 1 to 3 of lactation

Table 5.7. *Heritability (±SE) estimates etc. for yield test traits using model 1 with maternal lineage fitted or maternal lineage + herd * maternal lineage fitted.*

	Without maternal lineage		With Maternal Lineage		
Trait	h^2	F^2	h^2	f^2+hf^2	LRT
Milk test 1[a]	0.240 (0.028)	0.007 (0.006)	0.226 (0.028)	0.027 (0.009)	4.64
Fat test 1[a]	0.210 (0.024)	0.006 (0.006)	0.205 (0.027)	0.019 (0.009)	2.05
Protein test 1[a]	0.209 (0.026)	0.004 (0.006)	0.193 (0.025)	0.032 (0.008)	6.92
Milk test 2[a]	0.267 (0.028)	0.000 (0.000)	0.267 (0.028)	0.008 (0.008)	0.48
Fat test 2[a]	0.211 (0.018)	0.001 (0.005)	0.211 (0.021)	0.000 (0.000)	0.00
Protein test 2[a]	0.243 (0.025)	0.000 (0.000)	0.239 (0.022)	0.007 (0.007)	0.39
Milk test 3[a]	0.264 (0.028)	0.008 (0.006)	0.259 (0.029)	0.016 (0.008)	1.20
Fat test 3[a]	0.151 (0.018)	0.000 (0.000)	0.151 (0.018)	0.000 (0.000)	0.00
Protein test 3[a]	0.200 (0.025)	0.012 (0.006)	0.194 (0.025)	0.027 (0.009)	3.66

[a] Data set restricted to > 4 cows per maternal family

5.4 Discussion

The hypothesis that the inheritance of mtDNA has an effect on the difference in heritability estimates between daughter-dam and paternal half-sib estimation procedures in yield traits of dairy cattle is dependant on mtDNA having a significant effect on the traits in question. It is known that mitochondria are involved in the respiratory process and indeed are referred to as the 'energy factories' of cells. It has also been established that mitochondria contain a closed loop of 16,338bp of DNA. This mtDNA contains 12 protein-coding genes involved in energy production via the electron transport chain (Anderson *et al.*, 1982). Therefore the most likely involvement in production that follows from this is in highly energetic processes and times at which the cow is under metabolic stress. For this reason the yield traits in test days one to three around peak production were investigated. Previous studies have

looked at 305 day yield as the trait of interest (e.g. Boettcher and Gibson, 1997; Albuquerque *et al.*, 1998) which is understandable given that national evaluations are based upon this measure. In the present study no significant maternal lineage component of phenotypic variance was estimated for the 305 day yield traits. This is in agreement with all similar recent studies utilising an animal model for the estimation of maternal lineage variance (e.g. Boettcher and Gibson, 1997; Albuquerque *et al.*, 1998). The further investigation of test day yield for test days one to three also resulted in no significant lineage effects.

Persistency is another trait that is connected with energetic processes. It is desirable to have cows with flatter lactation curves to both allow a reduction in the metabolic stress of the cow and also to reduce the cost of feeding the cow (Sölkner and Fuchs, 1987). If the cow has a flatter lactation curve then more production energy can be obtained from roughage in the diet and concentrate costs can be reduced and flatter lactations make other management decisions easier, such as insemination decisions (Dekkers *et al.*, 1998). Sölkner and Fuchs (1987) suggested the use of SD3 due to the fact that it accounts for oscillations in the slope of the curve and is also on the same scale as milk yield. In this study a zero maternal lineage variance component was estimated for the SD3 trait, which is in contrast to the findings of Schnitzenlehner and Essl (1999) who estimated a 3.2% maternal lineage component for this measure of persistency in the Austrian Simmental population. However, persistency, as the ratio of milk production in day 1 to 100 of lactation over production in days 201 to 300 of lactation, was estimated to have a highly significant maternal lineage variance component of 4.4%. This is particularly substantial, when one considers that the heritability of this trait was estimated to be 10.4%. The component was only estimated when the data set was restricted to greater than 5 cows per maternal lineage, which was the same procedure as used by Schnitzenlehner and Essl (1999). This suggests that in the Holstein Friesian population maternal lineage has a significant effect on the maintenance of milk yield in the later part of the lactation curve. The

data editing procedure also resulted in a significant reduction in the heritability of SD3. The heritability estimates of persistency were both lower than those estimated by Sölkner and Fuchs (1987) in the Austrian Simmental population. They estimated SD3 and P3 heritability to be 0.19 and 0.21 respectively for first lactation. This compares with estimates for SD3 and P3 of 0.16 and 0.12 respectively when data was restricted to 2 or more cows per maternal lineage, and 0.12 and 0.14 when the restriction was 5 or more cows per maternal lineage. If a 4% component of phenotypic variance is attributable to maternal lineage variance for the persistency trait then, given the low heritability of this trait, there may be an impact on sire EBV. The impact of maternal lineage variance on the genetic gain in the US dairy industry was investigated by Boettcher *et al.*, (1996). They found that even a component as large as 10% would only have a minimal effect on genetic progress in milk yield based on a heritability of 0.25. However the estimated variance component for persistency was only 0.14. When maternal lineage was considered the heritability fell to 0.10 with the remaining 0.04 partitioned to maternal lineage variance. It is not unreasonable to hypothesise that when maternal lineage is responsible for such a large variance component in comparison to additive genetic variance there will be an overestimation of sire EBV when using sister and daughter information.

It is not easy to see why the restriction of data size to five or more cows per maternal lineage should have an effect on the estimation of maternal lineage variance for some traits. Small lineages make up the majority of cows that trace back to recent non-pedigree cows that have been graded-up by crossing. Given that these cows cannot be traced back to a distant cytoplasmic origin they may in fact be members of the same maternal family. The incorrect assignment of cows to different maternal lineages when they actually belong to the same lineage results in a reduction in the estimation of between maternal family lineage variance. This can be demonstrated by simulating a number of true maternal families then estimating between family variance with a structure that does not link all true family members together. This is the situation that

occurs when small maternal families are formed not tracing pedigrees back to true cytoplasmic origin. This is explained in Chapter 6.

The current study also attempted to look at preferential treatment (PT) of cow families. A discussion on preferential treatment was presented in Chapter 4, which explained that PT of cow families would not be detected using methods which look at within herd standard deviation of yield, and yet could have a significant effect on bull-dam selection. The hypothesis for cow family PT is that cows coming from known elite maternal lineages will receive PT in some herds. However this may not occur in all herds. In the current study about 3000 cow families were distributed over more than one herd and four families were distributed over more than twenty herds. In order to investigate PT of cow families a random effect was fitted in the model, which coded the family by herd effect. This effect was determined to account for both a family and family by herd interaction component. The model including this component was tested against the model that fitted maternal family in order to determine a significant interaction component. The only significant interaction component found was 3% for test day one protein yield. A component of 2.6% was also found for test day one milk yield but the test statistic of 4.64 would have been significant at the α=0.05 level if three or less independent tests were being carried out. It is not easy to see how PT can have an influence on protein yield in the first test day of the first lactation but it is possible to see how such an effect could occur in milk yield. The estimation of such a component could alternatively be an artifact of the data structure given that there are many small family*herd groups.

The analysis described was designed to provide a clear relationship structure to estimate maternal lineage variance if present. Contemporary field data was used to avoid the presence of maternal inheritance being incorrectly detected when using a maternal lineage component model. The early part of lactation was also investigated using test day records and no significant maternal lineage variance was estimated for

any of the yield traits. A maternal lineage variance component was, however, found for a measure of persistency and this should be considered when persistency is being evaluated.

Chapter 6

Bias and power in the estimation of a maternal family variance component in the presence of incomplete or incorrect pedigree information

6.1 Introduction

n the preceding chapters the magnitude of the between maternal lineage variance has been estimated for a variety of production and conformation traits in the UK Holstein Friesian population. The hypothesis is that this component of the overall phenotypic variance is attributable almost exclusively to the maternal transmission of mtDNA in mammals (Hutchinson *et al.*, 1974; Gyllensten *et al.*, 1991). It was shown that using an animal model to adjust for the additive nuclear genetic variance component enables one to estimate a simulated component of maternal lineage variance (Southwood *et al.*, 1989). This approach has been adopted by various studies over recent years (e.g. Boettcher *et al.*, 1996a; Schnitzenlehner and Essl, 1999), and was the approach taken in Chapters 2, 4 and 5. However this approach assumes that if variance exists between the true maternal lineages, i.e. families formed from the points of mtDNA divergence, then it is possible to estimate this variance component by tracing maternal lineages as described in Chapter 1. There are two main mechanisms involved by which there is likely to be a reduction in the magnitude of the maternal lineage variance component estimated. The first of these mechanisms is the incomplete assignment to true maternal families, i.e. the assignment of several maternal sub-families within one true larger family, which is not detected due to tracing insufficient generations in the establishment of complete maternal families. The second of these mechanisms is the incorrect assignment of pedigree leading to the accumulation of pedigree errors over generations from maternal lineage origin. The magnitude of these effects on the estimation of the maternal lineage variance

component is investigated. The effect that the two mechanisms have on the power to detect maternal lineage variance is also investigated.

6.2 Family Structure

The effect of tracing maternal lineages insufficient generations to establish the points of cytoplasmic origin can be illustrated by considering a simple balanced family design (Figure 6.1). In this design we assume that all cows in the current generation are an equal number of generations from their cytoplasmic origin and that all other fixed and random effects have been adjusted for.

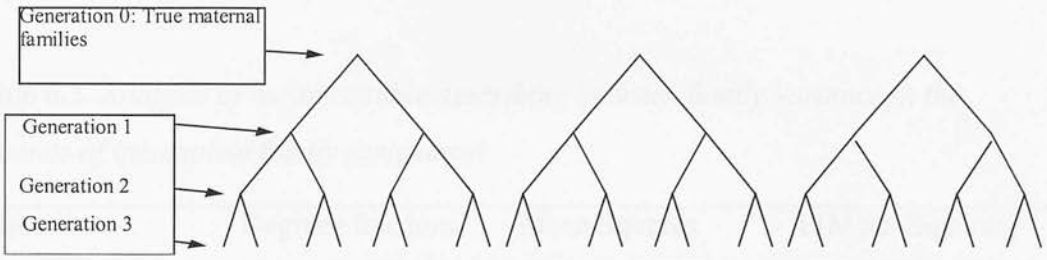

Figure 6.1. Illustration of section of family structure demonstrating sub families within true families

If full pedigree information is available, i.e. all records are assigned to the true maternal lineage, generation 0 in Figure 6.1, and the design is balanced, i.e. equal numbers per family, then the between family variance component ($\hat{\sigma}_f^2$) can be estimated using the analysis of variance (ANOVA) table illustrated in Table 6.1. If we then move to the situation where incomplete family information is available due to tracing insufficient generations to the origin, i.e. to generation 1, 2, or 3 in Figure 6.1, we move to the situation illustrated in Table 6.2. In this table there is now an extra level of sums of squares due to the between assigned sub-family replicate variance.

However, when performing statistical analyses with incomplete pedigree information, a combination of the between family and between replicate sums of squares is used, since the true structure is not recognised. In the situation where complete family assignment occurs (i.e. r = 1) Table 6.2 collapses to Table 6.1.

Table 6.1. *Analysis of variance table describing between family variance*

Source	Degrees freedom	Mean Squares	E[Mean Squares]
Between families	f-1	C	$(n)\sigma^2_f + \sigma^2_w$
Residual	f(n-1)	W	σ^2_w

f = number of true cow families

n = assumed number of cows per family

Table 6.2. *Analysis of variance table describing between family variance in the presence of incomplete family assignment*

Source	Degrees freedom	Mean Squares	E[Mean Squares]
Between families	f-1	B	$(nr)\sigma^2_f + \sigma^2_w$
Between replicates	f(r-1)	R	σ^2_w
Residual	fr(n-1)	W	σ^2_w

f = number of true cow families

r = number of replicates (i.e. number of sub families assigned to original family)

n = assumed number of cows per family

In order to obtain an estimate of the maternal family variance component we have an estimate of the sum of the sums of squares between replicates (f(r-1)R) and the sums of squares between true maternal lineages ((f-1)B). In practice we use the sum of the between family and between replicates sums of squares (equation 6.1).

$$SSC = SSB + SSR = (f-1)B + f(r-1)R \tag{6.1}$$

The SSC has (f-1) + f(r-1) = fr – 1 degrees of freedom because the apparent number of families is fr. The mean squares of this sum of two sums of squares is obtained by equation 6.2.

$$C = \frac{SSC}{fr-1} = \frac{[(f-1)B + f(r-1)R]}{fr-1} \tag{6.2}$$

The distribution of C is not proportional to a central χ^2. The expectation of the estimated maternal lineage variance $(\hat{\sigma}_f^2)$ is then given by equation 6.3.

$$E[\hat{\sigma}_f^2] = \frac{E[C-W]}{n} \tag{6.3}$$

It is shown in Equation 6.4 that the between lineage variance component is reduced as the number of replicates within assigned families is increased.

$$\frac{E[C-W]}{n} = \frac{1}{n}\left[\left(\frac{f-1}{fr-1}\right)\left(nr\sigma_f^2 + \sigma_w^2\right) + \frac{f(r-1)}{fr-1}\sigma_w^2 - \sigma_w^2\right]$$

$$= \sigma_f^2\left(\frac{(f-1)r}{fr-1}\right) + \sigma_w^2\left(\frac{(f-1)}{n(fr-1)} + \frac{f(r-1)}{n(fr-1)} - \frac{1}{n}\right)$$

$$= \sigma_f^2\left(\frac{(f-1)r}{fr-1}\right) + \sigma_w^2\left(\frac{fr-1+f-f}{n(fr-1)} - \frac{1}{n}\right)$$

$$= \sigma_f^2\left(\frac{(f-1)r}{fr-1}\right) \tag{6.4}$$

The downward bias introduced by incomplete maternal lineage assignment is illustrated in equation 6.5. The proportion of the maternal lineage variance component that is estimable after removing the bias demonstrated in Equation 6.5 is shown in Figure 6.2 for different family size and number of sub-families assigned within true family.

$$E\left[\hat{\sigma}_f^2\right] = \frac{(f-1)r}{fr-1} \cdot \sigma_f^2$$
$$\rightarrow Bias = \frac{\hat{\theta}-\theta}{\theta} = \frac{(f-1)r}{fr-1} - 1 = \frac{1-r}{fr-1} \qquad (6.5)$$

Example 6.1

Consider that the true number of families (f) is equal to 8. The total number of individuals, N is equal to 4096 and the current generation is 9 generations from the true origin. If at every generation the number of females is doubled, and the true between maternal lineage variance component is 5% of the phenotypic variance, then the magnitude of the variance component estimated for different numbers of generations traced is shown in Table 6.3.

Table 6.3. *Expectation of estimation of variance component by tracing different numbers of generations when the true maternal lineage variance component is 0.05*

Number of generations traced	Number of cows in assigned family	Number of assigned families	Variance component estimated
1	2	2048	0.0438
2	4	1024	0.0438
3	8	512	0.0438
4	16	256	0.0438
5	32	128	0.0441
6	64	64	0.0444
7	128	32	0.0452
8	256	16	0.0467
9	512	8	0.0500

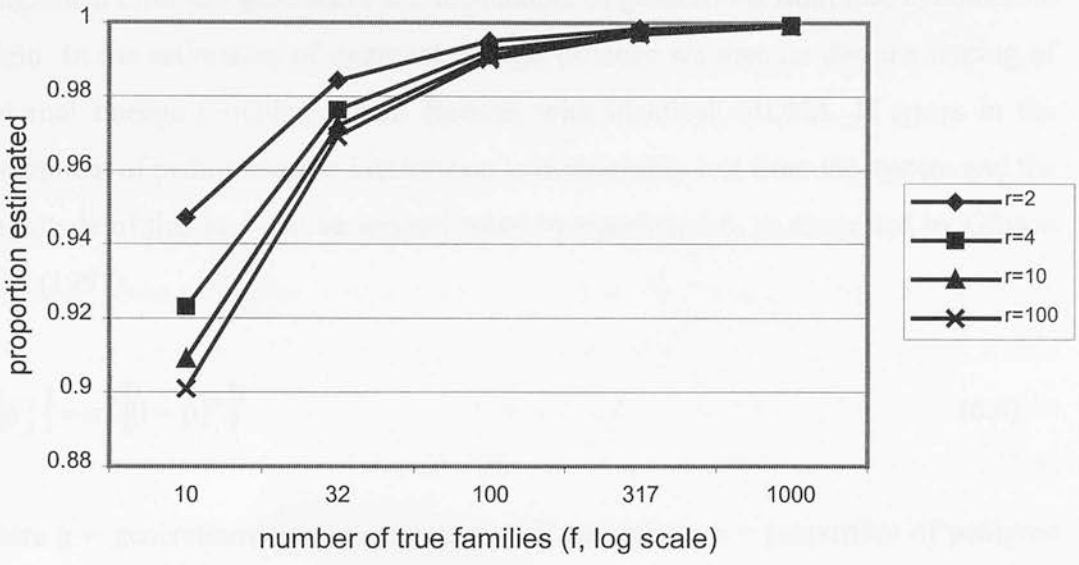

Figure 6.2. The proportion of the between lineage variance estimable with assignment to r sub-families within true family (f).

The downward bias introduced by not tracing families to their true cytoplasmic origins increases as the number of true families decreases, equation 6.5. For example the data set used in Chapter 5 had 30,000 records assigned to 10,000 families. If we assume that family size was equal in this case (though it was far from equal) and then we assume that there are either 1000, 100 or 10 true families then the downward bias is respectively 0.1%, 1% and 10%. It can clearly be seen that the impact is noticeable only if there are a small number of true families. However, in a modern dairy breed there could potentially be a low number of original mtDNA sources.

6.3 Pedigree Errors

Pedigree errors cause a build up of error in the estimation of maternal lineage variance, where the magnitude of under-estimation is affected by the level of

assignment error per generation and the number of generations from true cytoplasmic origin. In the estimation of maternal lineage variance we assume that the tracing of maternal lineage provides correct families with identical mtDNA. If errors in the assignment of pedigree occur information is irretrievably lost from the system and the magnitude of this loss can be approximated by equation 6.6, as suggested by Gibson *et al.* (1997).

$$E\left[\hat{\sigma}_f^2\right] = \sigma_f^2\left[(1-p)^g\right]^2 \qquad (6.6)$$

where g = generations from origin to current generation; p = proportion of pedigree errors per generation.

For example an error rate of 5% per generation results in a reduction in the magnitude of the estimated component of $[(1\text{-}0.05)^8]^2$ = 0.44 after 8 generations from cytoplasmic origin.

6.4 Combining the effect of incomplete and incorrect pedigree information

Equation 6.7 combines the effect of incomplete and incorrect pedigree information on the downward bias of maternal lineage variance estimation.

$$E\left[\hat{\sigma}_f^2\right] = \left(\frac{(f-1)r}{fr-1}\right)\cdot\left((1-p)^{2g}\right)\cdot\sigma_f^2$$
$$\rightarrow Bias = \frac{(f-1)r}{fr-1}\cdot(1-p)^{2g} - 1 \qquad (6.7)$$

6.5 Variance component estimation by simulation

To demonstrate the previous two effects simulations were run 10,000 times, each based on the structure shown in Table 6.2 and the family structure described in example 6.1, with the magnitude of the true maternal lineage variance and the error rate per generation being varied between simulations. Phenotypic records were simulated given a number of true maternal families. The pedigree error rate was then accounted for by calculating the proportion of the current generation that we would expect to be incorrectly assigned after the number of generation between the origin and the generation being simulated. The proportion of incorrect records expected were then randomly chosen and given a phenotype comprising only a random error component, assuming that the records are from random unknown maternal families. Using this phenotypic information a one way ANOVA was used to estimate the between family variance component with sub-division of the true simulated families to represent the incomplete tracing of pedigree information.

Equation 6.7 was able to predict the outcome of the simulation consistently to within ~5% of the true component. Two extreme examples of the results of this simulation are shown in tables 6.4 and 6.5. The pedigree error rate is 8% per generation and the records are 8 generations from the true cytoplasmic origin resulting in only $(1-0.08)^8 = 0.51$, i.e. 51% of records in the current generation being correctly assigned.

Table 6.4. *Simulated versus predicted magnitude of variance component where true variance component is 0.05 and pedigree error rate per generation is 0.08*

Number of families assigned	Number in assigned family	Predicted magnitude of variance component	Variance component estimated in simulation	predicted as a proportion of simulated variance component
2048	2	0.0115	0.0118	0.97
1024	4	0.0115	0.0116	0.99
512	8	0.0115	0.0115	1.00
256	16	0.0116	0.0116	1.00
128	32	0.0116	0.0117	0.99
64	64	0.0117	0.0118	0.99
32	128	0.0119	0.0120	0.99
16	256	0.0123	0.0124	0.99
8	512	0.0132	0.0132	1.00

Table 6.5. *Simulated versus predicted magnitude of variance component where true variance component is 0.5 and pedigree error rate per generation is 0.08*

Number of families assigned	Number in assigned family	Predicted magnitude of variance component	Variance component estimated in simulation	predicted as a proportion of simulated variance component
2048	2	0.1153	0.1145	1.01
1024	4	0.1153	0.1147	1.01
512	8	0.1155	0.1148	1.01
256	16	0.1157	0.1150	1.01
128	32	0.1161	0.1155	1.01
64	64	0.1171	0.1164	1.01
32	128	0.1190	0.1182	1.01
16	256	0.1229	0.1222	1.01
8	512	0.1317	0.1309	1.01

6.6 Power of detecting a maternal lineage component

When the incorrect assignment of pedigree is considered the power of detection of a variance component is the power of detection of the estimable component as obtained using equation 6.6. If complete pedigree information were available the variance ratio test statistic obtained in the presence of incorrect pedigree follows approximately an F distribution with the degrees of freedom (fr – 1) of the numerator and fr(n – 1) of the denominator of the ratio. However when the incomplete pedigree situation is considered it is not easy to ascertain the distribution of the test statistic. The test statistic obtained is the ratio of the mean squares C from Equation 6.2 over the residual mean squares. The mean squares C is derived from the sum of two sums of

squares (Equation 6.2) and is therefore not a standard χ^2. Using the simulation described in section 6.5 the power of detection of a 1% and 5% between lineage variance component was obtained (Table 6.6). The power was obtained comparing the ratio C/W to an F ratio test statistic with the degrees of freedom as previously described in this section to either accept or reject the variance component estimated at a 5% type-1 error rate in each simulation. Also presented in Table 6.6 is the power of detection of a between lineage variance component given that the assigned families are the true families. For this it was assumed that assigned families were true families and an F ratio power test was used again at a 5% level of type-1 error rate (e.g. Lynch and Walsh, 1998). It was seen that the actual power of detection was consistently lower than the situation where the number of families assigned were the true number of families. A further illustration of these results is shown in Figure 6.3 where the power is plotted for different magnitudes of true variance component for the situation where f*r true families exist and where each of the f families are assigned to r sub-families. No pedigree errors are included in the power results.

Table 6.6. *Power of detection of between lineage variance with incomplete pedigree information and no incorrect pedigree assignment*

Number of families assigned	Number in assigned family	Power of detecting variance component simulated		Predicted power if number of assigned families is true number families	
		1%	5%	1%	5%
2048	2	0.107	0.580	0.117	0.732
1024	4	0.179	0.802	0.195	0.982
512	8	0.288	0.915	0.322	1.000
256	16	0.430	0.961	0.510	1.000
128	32	0.578	0.983	0.723	1.000
64	64	0.719	0.993	0.876	1.000
32	128	0.831	0.998	0.942	1.000
16	256	0.901	0.998	0.958	1.000
8	512	0.941	1.000	0.941	1.000

6.7 Conclusions

It has been clearly demonstrated that there are two main mechanisms by which we can expect to under-estimate maternal lineage variance. The first mechanism is the result of tracing insufficient generations of the maternal pedigree, which to some extent can be overcome by a more detailed tracing procedure. For large data sets the feasibility of tracing of further generations of maternal lineage is restricted by the size of the pedigree already available on a database. For some situations to trace historic pedigrees further than the currently computerised pedigree is not practical. It was also seen that this is not a linear increase in the accuracy of estimation and is very

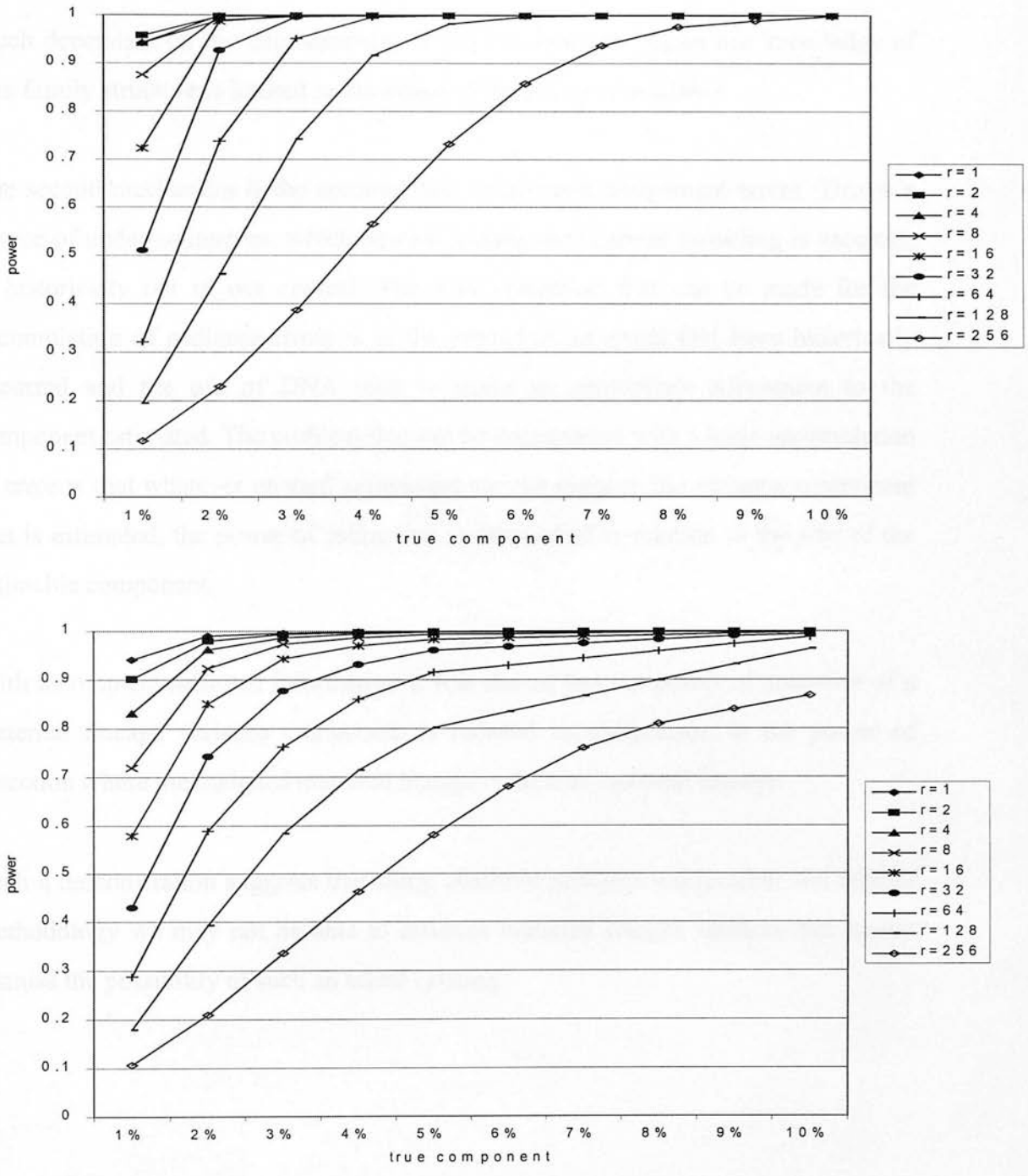

Figure 6.3. Power of detecting a true variance component if N=4096 at a 5% type-1 error rate. Top graph illustrates the power where assigned number (8*r) of families is true number of families. Bottom graph shows the simulation results when 8 true families (f) are incorrectly assigned to r sub-families.

much dependant on the data structure of the true families. Again our knowledge of true family structure is limited to the extent of the pedigree available.

The second mechanism is the accumulation of pedigree assignment errors. This is a source of under-estimation, which beyond making sure current recording is accurate, is historically out of our control. The only correction that can be made for the accumulation of pedigree errors is in the prediction of errors that have historically occurred and the use of DNA tests to make an appropriate adjustment to the component estimated. The problem that can be encountered with a large accumulation of error is that whatever upward adjustment we can make to the variance component that is estimated, the power of estimation is diminished in relation to the size of the estimable component.

With incomplete pedigree information it was shown that the power of detection of a maternal lineage variance component is reduced in comparison to the power of detection where the assigned maternal lineage is the true maternal lineage.

Such a demonstration suggests that using available pedigree information and current methodology we may not be able to estimate maternal lineage variance but cannot dismiss the possibility of such an effect existing.

Chapter 7

Conclusions and perspectives

7.1 Overview

In Chapter 1 the work previously undertaken to estimate the magnitude of maternal lineage variance was reviewed and the appropriate methodology to estimate such an effect was established. The animal model was shown to be able to give an unbiased estimate of maternal lineage variance using a simulation study (Southwood *et al.*, 1989). If a sire model is fitted (e.g. Bell *et al.*, 1985) an over-estimation of maternal lineage variance is obtained which could be due to selection and drift not being accounted for in the maternal lineage. In mammals the mtDNA is almost exclusively maternally inherited. The mtDNA is responsible for encoding some of the proteins involved in the respiratory process, the remainder being encoded by nuclear genes. This involvement in the energetic process is the basis used to explain the expression of maternal lineage effects. It is not unreasonable to assume that evidence to support such a hypothesis would be expected to be found for phenotypic traits that have a high energy demand.

In order to investigate energetic traits the extensively recorded Langhill dairy herd was used (Chapter 2). This allowed the investigation of feed intake and efficiency of milk production. No evidence of maternal lineage effects was found for either the feed intake or the efficiency of production traits. A maternal lineage variance component of 4%, that was close to a 5% type-1 error level of significance, for fat yield was estimated. An investigation of the maximum power of an analysis of a data set of the magnitude available from Langhill also revealed that a component of variance of less than 3% of the overall phenotypic variance would have a power of

detection of less than 70%. From this we can conclude that though no significant maternal lineage effects were detected for energetic traits it is not possible to state that they do not exist. A further explanation for not estimating a significant maternal lineage effect was identified by theoretical work in Chapter 6. Here it was demonstrated that if incorrect pedigree assignment occurred then under-estimation of maternal lineage effects would result. However the pedigree information that was used in Chapter 2 represented the maximum amount of information available for the UK Holstein Friesian population, i.e. all herd book information was used from the time the breed society was established. The historic accumulation of pedigree recording errors is still a potential source of under-estimation of maternal lineage effects.

Chapter 2 demonstrated that for yield traits maternal lineage effects are very unlikely to be of the magnitude originally estimated by the study of Bell *et al.* (1985). It was not possible to conduct an analysis of feed intake and the efficiency of production on a larger scale. Traits recorded in the UK dairy population enabled the investigation of aspects of conformation and production for the presence of maternal lineage effects.

Before undertaking a larger scale study, work was done to establish some population parameters of the UK dairy population (Chapter 3). This provided an insight into the nuclear genetic diversity of the population and also gave an indication of the maternal family structure. The analysis implemented in Chapters 4 and 5 for conformation and production traits respectively utilised contemporary groups of cows. Contemporary groups were used in order to minimise the average nuclear genetic relationship and the environmental covariance between cows within a maternal family. Thus the confounding of within family covariances with cytoplasmic inheritance was reduced. The contemporary group methodology in Chapters 4 and 5 necessitated the existence of large contemporary groups of cows in any single maternal family. In Chapter 3 the

structure of the cow families was shown to provide an acceptable data structure though approximately 40% of the records available occurred in maternal families of one cow with an ancestor and these records were discarded before the analysis.

Of the conformation traits only stature and the composite body trait were found to have a component of phenotypic variance attributable to maternal lineage effects. The magnitudes of the effects were 1.1% and 1.5% for stature and body respectively. Maternal lineage effects on production traits were not detected for the 305 day yield traits or the early lactation yield test day traits. However the test information on yield traits allowed the derivation of a trait describing the persistency of lactation between the first and last thirds of a 300 day lactation. For this derived trait a significant maternal lineage effect was found which accounted for 4.4% of the phenotypic variation of the trait. Given that this trait has a low heritability (0.1 to 0.3) such a component has implications in the use of bull persistency estimates based on daughter and female-sib information. When such an evaluation is done for bulls based on the performance of female relatives there is a potential overestimation of predicted transmitting ability. If a maternal lineage variance component exists and is ignored in the analysis then the heritability is over-estimated by an animal model or daughter-dam regression procedure. If the heritability that is used in an evaluation is inflated, too much emphasis is given to an animal's own performance. A dam could 'correctly' predict the performance of a daughter, but not a son. A simulation study by Boettcher *et al.* (1996c), based on the USDA (United States Department of Agriculture) evaluation procedure, showed that the bias introduced into sire PTA was not large. This is what would be expected because progeny testing of sires involves the use of information from many daughters. These daughters are a random sample of the maternal lineages in the population and therefore any maternal lineage effects will be averaged out and the ranking of bulls will not be affected. The overall population result of the simulation estimated a change in the correlation between true breeding value and predicted breeding value of 0.009 for a maternal lineage component of

2.5% and 0.042 for a component of 10%. It was also shown by simulation that if there were a maternal lineage effect then accounting for this in bull evaluations could allow a reduction in the size of progeny test schemes, leading to substantial savings. One area that has not been investigated, probably due to a low perceived selection intensity, is the potential improvement by selection of daughters from superior maternal lineages. The intensity of this selection pathway can now be substantially increased through the use of reproductive technology used in MOET dairy breeding schemes. However, in a MOET scheme contemporary, i.e. sister information is used to calculate sire EBV and here there is potentially a severe bias being introduced. The distinction here is that unlike in progeny testing where the information used is averaged across many maternal lineages, in a MOET scheme the information used to evaluate a bull comes from only a few maternal lineages. This potentially introduces a major bias into the evaluation and if we consider a trait such as persistency with a low heritability and high maternal lineage variance then a sire's estimated breeding value (EBV) for persistency would be heavily influenced by the maternal lineage effect of his full and half sisters.

In Chapters 4 and 5 the tracing of cow families for the large data set was only possible using those pedigree records that are stored on the HUKI computer database, which holds registrations from around 1960 onwards. If the results of Chapters 4 and 5 are viewed in the context of the potential under-estimation described by Chapter 6, where the incorrect assignment of cows to sub-families within true families occurs, then it is again not possible to say that maternal lineage effects do not affect phenotypic traits. However the methodology employed was not able to estimate such effects in the absence of complete maternal family information.

7.2 Future Perspective

It was demonstrated (Kennedy, 1986) that the original estimates of maternal lineage effects (e.g. Bell *et al.*, 1985; Huizinga *et al.*, 1986) were likely to be inflated due to inadequate models. It was also shown by simulation that it is possible to detect a simulated maternal lineage variance component using an animal model (Southwood *et al.*, 1989). However in Chapters 2, 4 and 5 only very few traits were identified as having significant maternal lineage effects affecting their phenotype. The magnitude of maternal lineage effect, though likely to be smaller than originally estimated, could still exist in the light of the potential under-estimation highlighted by Chapter 6. If we accept that current methodology is unlikely to be capable of identifying a small maternal lineage variance component then we must look to alternative methods.

In the new biological technology age it becomes more important than ever to fully understand how the genetic material of an organism contributes to its phenotype. Indeed in the light of cloning and nuclear transfer technology a full understanding of the contribution of mtDNA and its interaction with nuclear DNA becomes important. During the cloning process nuclear genetic material is transferred. This process occurs by the fusion of a donor somatic cell with a recipient oocyte using electroportation. The expectation from this process is that the clone habours mtDNA from both donor and recipient sources. However it was recently shown that when a cloned sheep was produced using this process, it was the mtDNA of the recipient only that was present in the cloned animal (Evans *et al.*, 1999). This illustrates how important it is to understand both the contribution that mtDNA makes and also the interaction that occurs with nuclear DNA.

In Chapter 1 experimental design was described (Beavis *et al.*, 1987) that would separate maternal lineage variance accurately from other sources of variation with

only one confounding interaction. In combination with the use of reproductive technology such an experiment is not impossible in dairy cattle but in order to achieve the size of experiment required it is unlikely to be feasible. The approach required would appear to be a combination of current and developing methods. If the only information that is available is pedigree and phenotypic records then certain limitations must be considered. As highlighted in Chapter 6 the issues of incorrect and incomplete pedigree causes a downward bias and while there is little that can be done about historic pedigree error one should attempt to construct the most complete pedigree information that is available and feasible. The power of the analysis to detect maternal lineage variance depends upon the size of the data set, the magnitude of the component and the maternal lineage structure. It was shown in Chapter 2 that with only 500 records in a balanced design with accurate lineage information the power of detecting a 3% variance component would be less than 70%. However if we move to 1000 records with the same structure the power of detection is over 90%. It was demonstrated in Chapter 6 that the power of detection is potentially further diminished because of possible incomplete or incorrect pedigree assignment. This is dependent upon the underlying true family structure, which we cannot know beyond the pedigree information available. It must also be remembered that these studies on the power of maternal lineage variance detection were based on balanced family information and therefore represent the upper limit of detection.

With the use of only phenotypic and pedigree information it is not to validate that the hypothesis proposed is true, i.e. the maternal inheritance of mtDNA is the correct one. There are a number of alternative hypotheses as mentioned in Chapter 1 such as maternal effects, sex-linkage and imprinting. In a controlled experiment these effects all display unique patterns of inheritance but in the data structures that have been used in the analysis of dairy cattle it may not be possible to separate their effects. Even in carefully designed experiments Meyer (1992) showed that estimates of maternal effects and pertaining variance components are very imprecise.

With the rapid expansion in the availability of molecular technology comes the gradual reduction in the cost of the process. At the present time the search for specific polymorphisms connected to the encoding of proteins involved in the respiratory process in mtDNA is not economically feasible. Work has been done looking at the hypervariable D-loop region of mtDNA (e.g. Schutz *et al.*, 1994; Ron *et al.* 1993) for polymorphisms but as highlighted in Chapter 1 this is not a protein coding region and any detected effect is by association (not causal). More extensive sequencing techniques are very likely to become economically feasible in the future and in combination with appropriate statistical procedures, help with the identification of advantageous sources of mtDNA. If databases based on either complete mtDNA sequence, or more realistically marker information are developed then problems of maternal family assignment can be overcome and the assumption that mtDNA is identical within these families can be verified. Techniques such as these using nuclear DNA are already used in the establishment of paternity, though at present the cost is prohibitive to a large scale mtDNA study. Improvement could then be facilitated through this identification combined with the rapid dissemination of such mtDNA either through the future use of cloning or by techniques currently used such as multiple ovulation embryo transfer.

There are several approaches to the design of an experiment utilizing molecular technology. Initial approaches have used phenotypic information to identify the highest and lowest producing maternal families (e.g. Ron *et al.*, 1993). This reliance on initial phenotypic family information to perform selective genotyping may not be the optimal approach given the potential historical accumulation of pedigree errors. Another approach is to take a random sample from the population for genotyping and to look at sequence variation within and between maternal family based on pedigree information. The next step would very much depend on the result of this initial step. If it were shown that within lineage mtDNA sequence is uniform or that a pedigree

error rate can be established, then the use of a more conventional animal model approach using pedigree information is an adequate method. Such a method still suffers from the problem of competing genetic hypotheses as discussed previously in this section. If however large within lineage sequence variation is identified then the use of pedigree information to define maternal lineage is inadequate. One method given this scenario is to use sequence information to define maternal lineage in a refinement of previous molecular approaches which used only D-loop variation (e.g. Schutz *et al.*, 1993), and to incorporate this information into a more traditional pedigree based animal model approach. This method would not rely on long term accurate pedigree records and in fact we may conceivably be able to construct or at least verify pedigree information based upon genotype information.

Given the underlying assumption that mtDNA within a maternal lineage is identical in all cows, then if advantageous sources of mtDNA are identified and herds are established based on such lineages this is a one off process and there is no potential for continuous improvement based on mtDNA, apart from new mutation.

References

Albuquerque, L.G., Keown, J.F., VanVleck, L.D. 1998. Variances of direct genetic effects, maternal genetic effects, and cytoplasmic inheritance effects for milk yield, fat yield, and fat percentage. *Journal of Dairy Science*. **81**:544-549.

Anderson, S., deBruijn, M.H.L., Coulson, A.R., Eperon, I.C., Sanger, F., Young, I.G. 1982. Complete Sequence of Bovine Mitochondrial DNA. Conserved Features of the Mammalian Mitochondrial Genome. *Journal of Molecular Biology*. **156**:683-717.

Beavis, W.D., Pollak, E. and Frey, K.J. 1987. A theoretical model for quantitatively inherited traits influenced by nuclear-cytoplasmic interactions. *Theoretical and Applied Genetics*. **74**:5711-578.

Bell, B.R., McDaniel, B.T., Robinson, O.W. 1985. Effects of cytoplasmic inheritance on production traits of dairy cattle. *Journal of Dairy Science*.**68**:2038-2051.

Bentler, P.M. and Yuan, K., 1996. Test of linear trend in eigenvalues of a covariance matrix with application to data analysis. *British Journal of Mathematical and Statistical Psychology*, **49**:299-312.

Boettcher, P.J., Freeman, A.E., Johnston, S.D., Smith, R.K., Beitz, D.C., McDaniel, B.T. 1996a. Relationships between polymorphism for mitochondrial deoxyribonucleic acid and yield traits of Holstein cows. *Journal of Dairy Science*. **79**:647-654.

Boettcher, P.J., Steverink, D.W.B., Beitz, D.C., Freeman, A.E., McDaniel, B.T. 1996b. Multiple herd evaluation of the effects of maternal lineage on yield traits of Holstein cattle. *Journal of Dairy Science*. **79**:655-662.

Boettcher, P.J., Kuhn, M.T., Freeman, A.E. 1996c. Impacts of cytoplasmic inheritance on genetic evaluations. *Journal of Dairy Science*. **79**:663-675.

Boettcher, P.J., Gibson, J.P. 1997. Estimation of variance of maternal lineage effects among Canadian Holsteins. *Journal of Dairy Science*. **80**:2167-2176.

Boichard, D., Maignel, L., Verrier, E. 1997. The value of using probabilities of gene origin to measure genetic variability in a population. *Genetics Selection Evolution* **29:**5-23.

Bowman, J.C., Butler, E.A., Tuncel, E. 1978. Coefficients of inbreeding and degree of relationship for the British Friesian herd. *Animal Production* **27:**269-276.

Brotherstone, S., McManus, C.M. and Hill, W.G., 1990. Estimation of genetic parameters for linear and miscellaneous type traits in Holstein-Friesian dairy cattle. *Livestock Production Science*, **26**:177-192.

Brotherstone, S. and Hill, W.G., 1991. Dairy herd life in relation to linear type traits and production 2. Genetic analyses for pedigree and non-pedigree cows. *Animal Production,* **53**:289-297.

Brown, D.R., Koehler, C.M., Lindberg, G.L., Freeman, A.E., Mayfield, J.E., Myers, A.M., Schutz, M.M., Beitz, D.C. 1989. Molecular analysis of cytoplasmic genetic-variation in Holstein cows. *Journal of Animal Science.* **67**:1926-1932.

Brown, D.R., Denise, S.K., McDaniel, R.G. 1988. Mitochondrial respiratory metabolism and performance of cattle. *Journal of Animal Science.* **66**:1347-1354.

Brown, D.R., Denise, S.K., McDaniel, R.G. 1989. Cytoplasmic genetic-effects and growth of hybrid mice. *Journal of Animal Science.* **67**:887-894.

Brown, D.R., Denise, S.K., McDaniel, R.G. 1987. Mitochondrial respiratory metabolism and performance of Holstein cows. *Federation Proceedings.* **46**:1175.

Cattell, R.B., 1966. The scree test for the number of factors. *Multivariate Behavioural Research.* **1**:245-276.

Cockerham, C.C. 1967. Group inbreeding and coancestry. *Genetics* **56:**89-104.

Dekkers, J.C.M., Ten Hag, J.H., Weersink, A. 1998. Economic aspects of persistency of lactation in dairy cattle. *Livestock Production Science.* **53**:237-252.

Evans, M.J., Gurer, C., Loike, J.D., Wilmut, I., Schnieke, A.E., Schon, E.A. 1999. Mitochondrial DNA genotypes in nuclear transfer-derived cloned sheep. *Nature Genetics.* **23**:90-93.

Freeman, A.E. (1990). Cytoplasmic inheritance associated with economic traits - phenotypic and molecular differences. *Proceedings of the 4th World Conference on Genetic Applied to Livestock Production.* **XIV:**140-143.

Gibson, J.P., Freeman, A.E., Boettcher, P.J. 1997. Cytoplasmic and mitochondrial inheritance of economic traits in dairy cattle. *Livestock Production Science.* **47**:115-124.

Goddard, M.E. 1992. Optimal effective population size for the global population of black and white dairy cattle. *Journal of Dairy Science* **75:**2902-2911.

Graham, N.J., Smith, C. and Gibson, J.P. 1991. Investigation of preferential treatment for milk yield in Canadian Holsteins. *Canadian Journal Animal Science*. **71**:21-27.

Groeneveld, E. 1996. REML VCE a multivariate multi model restricted maximum likelihood (co)variance component estimation package version 3.2 user's guide. Federal Research Center of Agriculture, Mariensee, Germany.

Gyllensten, U., Wharton, D., Josefsson, A.,Wilson, A.C. 1991. Paternal inheritance of mitochondrial DNA in mice. *Nature* **352**:255.

Henderson, C.R. 1976. A simple method for computing the inverse of a numerator relationship matrix used in the prediction of breeding values. *Biometrics* **32:**69-83.

Hutchinson, C.A., Newbold, J.E., Potter, S.S., Edgell, M.H. 1974. Maternal inheritance of mammalian mitochondrial DNA. *Nature* **251**:536-538.

Huizinga, H.A., Korver, S., McDaniel, B.T., Politiek, R.D. 1986. Maternal Effects Due to Cytoplasmic Inheritance in Dairy Cattle. Influence on Milk Production and Reproduction Traits. *Livestock Production Science*. **15**:11-25.

Kaiser, H.F., 1958. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*. **23**:187-200.

Kennedy, B.W. 1986. A Further Look at Evidence for Cytoplasmic Inheritance of Production Traits in Dairy Cattle. *Journal of Dairy Science*. **69**:3100-3105.

Kirkpatrick, B.W., Dentine, R. 1988. An alternative model for additive and cytoplasmic genetic and maternal effects on lactation. *Journal of Dairy Science*. **71**:2502-2507.

Kuhn, M.T., Boettcher, P.J., Freeman, A.E. 1994. Potential biases in predicted transmitting abilities of females from preferential treatment. *Journal of Dairy Science*. **77**:2428-2437.

Kuhn, M.T. and Freeman, A.E. 1995. Biases in predicted transmitting abilities of sires when daughters receive preferential treatment. *Journal Dairy Science*. **78**:2067-2072.

Lacy, R.C. 1989. Analysis of founder representation in pedigrees: founder equivalents and founder genome equivalents. *Zoo Biology* **8:**111-123.

Lynch, M., Walsh, B. 1998. Genetics and analysis of quantitative traits. pp.885. Sinauer Associates, Inc., Sunderland, Massachusetts.

Meuwissen, T.H.E., Luo, Z. 1992. Computing inbreeding coefficients in large populations. *Genetics Selection Evolution* **24:**305-312.

Meyer, K. 1992. Bias and sampling covariances of estimates of variance components due to maternal effects. *Genetics Selection Evolution*. **24:**487-509.

Meyer, K., Brotherstone, S., Hill, W.G. and Edwards, M.R., 1987. Inheritance of linear type traits in dairy cattle and correlations with milk production. *Animal Science*, **44**:1-10.

Miglior, F., Szkotnicki, B., Burnside, E.B. 1992. Analysis of levels of inbreeding and inbreeding depression in Jersey cattle. *Journal of Dairy Science* **75:**1112-1118.

Miglior, F., Burnside, E.B. 1995. Inbreeding of Canadian Holstein cattle. *Journal of Dairy Science* **78:**1163-1167.

Olori, V.E., Brotherstone, S., Hill, W.G., McGuirk, B.J. 1999. Fit of standard models to the lactation curve to weekly records of milk production of cows in a single herd. *Livestock Production Science*. **58**:55-63.

O'Neill, K., Van Vleck, L.D. 1988. Potential of cytoplasmic effects for selection in dairy-cattle. *Journal of Dairy Science*. **71**:3390-3398.

Pander, B.L., Hill, W.G., Thompson, R. 1992. Genetic parameters of test day records of British Holstein-Friesian heifers. *Animal Production* **55**:11-21.

Pryce, J.E., Esslemont, R.J., Thompson, R., Veerkamp, R.F., Kossaibati, M.A., Simm, G. 1998. Estimation of genetic parameters using health, fertility and production data from a management recording system for dairy cattle. *Animal Science* **66**:577-584.

Reed, P.D., Van Vleck, L.D. 1987. Lack of evidence of cytoplasmic inheritance in milk-production traits of dairy cattle. *Journal of Dairy Science*. **70**:837-841.

Robertson, A. and Asker, A.A. 1951. The genetic history and breed structure of British Friesian cattle. *Empire Journal of Experimental Agriculture.* **19:**113-130.

Ron, M., Yoffe, O. and Weller, J.I. 1993. Sequence variation in D-loop mtDNA of cow lineages selected for high and low maternal effects on milk production. *Animal Genetics*. **24:**183-186.

Rorato, P.R.N., Keown, J.F., VanVleck, L.D. 1999. Variance caused by cytoplasmic line and sire by herd interaction effects for milk yield considering estimation bias. *Journal Of Dairy Science.* **82:**1574-1580.

Roughsedge, T., Brotherstone, S., Visscher, P.M. 1998. Lack of evidence for cytoplasmic inheritance for milk production traits at the Langhill dairy herd. *Proceedings of the sixth world congress on genetics applied to livestock production, Armidale.* **23**:351-354.

Salehi, A., James, J.W. 1997. Detection of cytoplasmic effects on production: The influence of number of years of data. *Genetics Selection Evolution.* **29**:269-277.

Schnitzenlehner, S. and Essl, A. 1999. Field data analysis of cytoplasmic inheritance of dairy and fitness-related traits in dairy cattle. *Animal Science.* **68**:459-466.

Schutz, M.M., Freeman, A.E., Lindberg, G.L., Koehler, C.M., Beitz, D.C. 1994. The effect of mitochondrial-DNA on milk-production and health of dairy cattle. *Livestock Production Science.* **37**:283-295.

Schutz, M.M., Freeman, A.E., Lindberg, G.L., Beitz, D.C. 1993. Effects of maternal lineages grouped by mitochondrial genotypes on milk-yield and composition. *Journal of Dairy Science.* **76**:621-629.

Schutz, M.M., Freeman, A.E., Beitz, D.C., Mayfield, J.E. 1992. The importance of maternal lineage on milk-yield traits of dairy cattle. *Journal of Dairy Science.* **75**:1331-1341.

Seykora, A.J. and McDaniel, B.T. 1983. Heritabilities and Correlations of Lactation Yields and Fertility for Holsteins. *Journal of Dairy Science.* **66**:1486-1493.

Simm, G., Persaud, P., Neilson, D.R., Parkinson, H., McGuirk, B.J. 1991. Predicting food intake in dairy heifers from early lactation records. *Animal Production* **52:**421-434.

Sölkner, J. and Fuchs, W. 1987. A comparison of different measures of persistency with special respect to variation of test-day milk yields. *Livestock Production Science.* **16**:305-319.

Sölkner, J., Filipcic, L., Hampshire, N. 1998. Genetic variability of populations and similarity of subpopulations in Austrian cattle breeds determined by analysis of pedigrees. *Animal Science* **67:**249-256.

Southwood, O.I., Kennedy, B.W., Meyer, K., Gibson, J.P. 1989. Estimation of additive maternal and cytoplasmic genetic variances in animal models. *Journal of Dairy Science*. **72**:3006-3012.

Stram, D.O. and Lee, J.W. 1994. Variance-components testing in the longitudinal mixed effects model. *Biometrics*. **50:**1171-1177.

Te Braake, M.F.H., Groen, A.F., Van Der Lugt, A.W. 1994. Trends in inbreeding in Dutch Black and White dairy cattle. *Journal of Animal Breeding and Genetics* **111:**356-366.

VanRaden, P.M. 1992. Accounting for inbreeding and crossbreeding in genetic evaluation of large populations. *Journal of Dairy Science* **75:**3136-3144.

Veerkamp, R.F., Simm, G., Oldham, J.D. 1994. Effect of interaction between genotype and feeding system on milk production, feed intake, efficiency and body tissue mobilization in dairy cows. *Livestock Production Science* **39:**229-241.

Visscher, P.M. and Thompson, R. 1992. Comparison between genetic variances estimated from different types of relatives in dairy cattle. *Animal Production* **55**:315-320.

Vukasinovic, N., Moll, J. and Künzi, N. 1997. Factor analysis for evaluating relationships between herd life and type traits in Swiss Brown cattle. *Livestock Production Science*, **49**:227-234.

Wagner, R.P. 1972. The role of maternal effects in animal breeding: II Mitochondria and animal inheritance. *Journal of Animal Science*. **35**:1280-1287.

Wang, J. 1997. More efficient breeding systems for controlling inbreeding and effective size in animal populations. *Heredity* **79:**591-599.

Weigel, D.J., Pearson, R.E. and Hoeschele, I. 1994. Impact of different strategies and amounts of preferential treatment on various methods of bull-dam selection. *Journal Dairy Science*, **77**:3163-3173.

Westhusin, M.E., Azambuja, R.M. 1996. Development of in vitro derived bovine embryos following pronuclear transplantation and in vitro culture. *Animal Reproduction Science*. **45**:29-35.

White, I.M.S., Thompson, R. and Brotherstone, S. 1999. Genetic and environmental smoothing of lactation curves with cubic splines. *Journal of Dairy Science*. **82**:632-638.

Wiggans, G.R., Van Raden, P.M. and Zuurbier, J. 1995. Calculation and use of inbreeding coefficients for genetic evaluation of United States Dairy Cattle. *Journal of Dairy Science* **78:**1584-1590.

Wilmink, J.M.B. 1987. Adjustment of test-day milk, fat and protein yield for age, season and stage of lactation. *Livestock Production Science.* **16**:335-348.

Wilmut, I., Schnieke, A.E., McWhir, J., Kind, A.J., Campbell, K.H.S. 1997. Viable Offspring Derived From Fetal and Adult Mammalian Cells. *Nature* **385**:810-813.

Woolliams, J.A., Mantysaari, E.A. 1995. Genetic contributions of Finnish Ayrshire bulls over four generations. *Animal Science* **61:**177-187.

Wray, N.R., Thompson, R. 1990. Prediction of rates of inbreeding in selected populations. *Genetical Research* **55:**41-54.

Wray, N.R., Woolliams, J.A., Thompson, R. 1994. Prediction of rates of inbreeding in populations undergoing index selection. *Theoretical and Applied Genetics.* **87:**878-892.

Wright, S. 1931. Evolution in Mendelian populations. *Genetics.* **16:**97-159.

Wright, S., McPhee, H.C. 1925. An approximate method of calculating coefficients of inbreeding and relationship from livestock pedigrees. *Journal of Agricultural Research* **31:**377-383.

Young, C.W. and Seykora, A.J. 1996. Estimates of inbreeding and relationship among registered Holstein females in the United States. *Journal of Dairy Science* **79:**502-505.

APPENDIX

Published papers :

Roughsedge T., Brotherstone S., Visscher P.M. (1998). Lack of evidence for cytoplasmic inheritance for milk production traits at the Langhill dairy herd. *Proceedings of the 6th world congress on genetics applied to livestock production, Armidale.* **23**: 351-354.

Roughsedge T., Brotherstone S., Visscher P.M. (1999). Estimation of variance of maternal lineage effects at the Langhill dairy herd. *Animal Science* **68**: 79-86.

Roughsedge T., Brotherstone S., Visscher P.M. (1999). Using pedigree analysis to determine the genetic diversity of the British dairy population over the last thirty years. *BSAS winter meeting 1999.*

Roughsedge T., Brotherstone S., Visscher P.M. (1999). Quantifying genetic contributions to a dairy cattle population using pedigree analysis. *Livestock Production Science.* **60**:359-369.

LACK OF EVIDENCE FOR CYTOPLASMIC INHERITANCE FOR MILK PRODUCTION TRAITS AT THE LANGHILL DAIRY HERD

T.Roughsedge , S. Brotherstone , P.M.Visscher

University of Edinburgh, West Mains Road, Edinburgh EH9 3JG

SUMMARY

Data on *717* first lactation cows at the Langhill Dairy Herd in Edinburgh, Scotland, were used to investigate the presence of a variance component attributable to cytoplasmic lineage. The cows , all Holstein Friesian, calved between 1979 and 1996 and were part of a long term selection experiment . Analysis was performed using an animal model and either fitting lineage, i.e. cow family, as a random variable or not including it. The analysis failed to provide any evidence of a significant variance component that could be attributed to cytoplasmic lineage, for milk, fat and protein yields, fat and protein percent or log transformations of the traits.

Keywords: cytoplasmic inheritance, milk yield, Holstein Friesian, dairy cattle, cow families

INTRODUCTION

Over recent years a great deal of interest has been expressed in the possible contribution of cytoplasmic inheritance to the inheritance of milk production traits in dairy cattle. The most likely cause of this relationship is the inheritance of mitochondrial DNA (mtDNA) which is additional to nuclear genetic inheritance (Gibson *et al.* , 1997). The reason for the interest in mtDNA is its exclusively maternal transmission in mammalian species (Hutchinson et al. , 1974). This might provide a possible explanation for the estimation of heritability being higher from daughter-dam regression than from paternal half sib correlations, (e.g. Seykora *et al. , 1983).*

Mitochondrial DNA only encodes about 0.05% of the total nuclear plus mitochondrial genes but the nature of the genes encoded make this a potentially important contribution. Mitochondria are responsible for a large part of the energy production of cells and as such it can be assumed that they will have a substantial effect on energetic processes such as the production of milk. The aim of this study was to investigate the presence of a cytoplasmic component of variance attributable to the cow families at the Langhill dairy herd for milk yield traits. The Langhill herd has been extensively recorded and a long term selection experiment has been in place there since the late seventies.

MATERIALS AND METHODS

Data description. Data were first lactation production records of 717 cows at the Langhill herd which calved between 1979 and 1996. The traits used for analysis were 305 day rnilk yield (MILK), fat yield (FAT), protein yield (PRT), fat% (FATP), protein% (PRTP), log milk yield (LMILK), log fat yield (LFAT) and log protein yield (LPRT). The reason for the inclusion of log transformations of the traits was intended to reduce the heterogeneous variance over time.

All cows were traced back to founder ancestors in the Langhill herd. These founder ancestors were then traced using the records of the Holstein Friesian Society to either the first point of registration of a founder female or to a cut off point of the year 1920, given no further convergence looked probable. The earliest cows traced were taken as being the points of cytoplasmic origin for the Langhill herd. This tracing resulted in the 717 first lactation cows being assigned to 78 cytoplasmic lineages (maternal lineages). The size of these groups ranged from the largest being 117 cows down to one.

Data analysis. Data were analysed with VCE (Groeneveld , 1996) using an animal model. Full and reduced models were fitted. The full model contained a cow lineage effect.

$$Y_{ijklm} = L_i + M_j + YS_k + b_1(age) + b_2(age^2) + a_l + f_m + e_{ijklm}$$

where Y_{ijklm} = MILK, FAT, PRT, FATP, PRTP, LMILK, LFAT, LPRT ; L_i = fixed effect of selection line ,(i is selected or control line) ; M_j = fixed effect of month of calving, (j=1, 12) ; YS_k = fixed effect of year-season of calving , (season was divided into three four month periods from Jan to April, May to Aug, Sept to Dec) ; age= age at calving; $b_1(age)$ = linear regression of Y on age ; $b_2(age^2)$ = quadratic regression of Y on age ; a_l = additive genetic effect of animal ; f_m = random effect of maternal lineage (m=1,78) ; e_{ijklm} = residual error. In the reduced model the random effect of maternal lineage (f_m) was removed.

RESULTS AND DISCUSSION

Overall means and standard deviations for the traits analysed for the 717 cows are in Table 1.

The results of the analysis of traits of first lactation cows at Langhill (Table 2) failed to demonstrate any detectable component of variance attributable to cytoplasmic line based on standard errors (from VCE) or a log likelihood ratio test (under the null hypothesis of no variation due to maternal lineage, the asymptotic distribution of the log likelihood ratio test is $^1/_2\chi^2(0) + {}^1/_2\chi^2(1)$, giving a 5% significance threshold of 2.7). It can however be seen that the standard errors associated with the components attributable to cytoplasmic lineage are equal

in magnitude to the components themselves and therefore we cannot rule out the presence of an effect.

The analysis corrected for the herd being composed of a selection and a control line by fitting the line as a fixed effect in the models. There was however the drawback, as with many experimental herds, of not being able to account for all the special management practices.

Table 1. Yield traits for 1st lactation records of 717 cows in 78 maternal lineages

	Untransformed		Log Scale	
Trait	Mean	SD	Mean	SD
Milk (kg)	6523	1255	8.76	0.19
Fat (kg)	274	52	5.59	0.19
Protein (kg)	213	39	5.34	0.19
Fat (%)	4.23	0.43		
Protein (%)	3.28	0.25		

Table 2. Heritability (±SE) estimates for the reduced model and the full (maternal lineage) model

	REDUCED MODEL	FULL MODEL		
Trait	h^2	h^2	f^2	LRT
MILK	0.32 (0.06)	0.28 (0.06)	0.03 (0.02)	1.62
FAT	0.37 (0.06)	0.35 (0.06)	0.01 (0.01)	0.54
PRT	0.25 (0.06)	0.23 (0.06)	0.01 (0.01)	0.44
FATP	0.64 (0.05)	0.60 (0.07)	0.02 (0.02)	0.48
PRTP	0.5 1 (0.06)	0.5 1 (0.06)	0.00 (0.00)	0.00
LMILK	0.3 1 (0.06)	0.27 (0.06)	0.03 (0.02)	1.52
LFAT	0.36 (0.06)	0.34 (0.06)	0.01 (0.01)	0.38
LPRT	0.23 (0.06)	0.21 (0.06)	0.01 (0.01)	0.48

f^2 is the proportion of phenotypic variance attributable to maternal lineage

LRT is twice the difference in log-likelihood between the full and reduced model

Mitochondrial inheritance would be expected to be responsible for a larger component of the traits highly dependent upon energetic processes. We would therefore expect traits such as fat yield or content to have a more significant variance component attributable to cytoplasmic lineage. The analysis did not provide significant evidence to support this theory. It was noticeable that protein% showed no variance component attributable to maternal lineage and fat% showed a larger maternal lineage component. If there was significant evidence for a maternal lineage component this is the trend we would expect.

The analysis was of data collected over seventeen years and there was a concern that the analysis would not be accounting for the change in variance over time within a selected herd. In order to correct in some way for this, log transformations of yield data were also analysed by the model. It was evident from the analysis that the transformed data provided the same results (Table 2) as the first analysis. Another- concern was the effect of the use of selected bulls on the analysis. This selection effect is not properley taken into account, even when fitting an animal model, because selection took place in the national herd. We plan to investigate this either by fitting sire as a fixed effect, or regressing the data on the sire estimated breeding values.

The Eanghill herd is extensively recorded for traits such as feed intake and the investigation will now move to look at traits such as production efficiency which are heavily energy dependent and therefore potentially dependent on mitochondria.

ACKNOWLEDGMENTS

T Roughsedge is supported by MAFF. We would like to thank Roel Veerkamp and Ross McGinn for providing data, and all staff at Langhill for their excellent work. Thanks to Sara Knott for helpful comments.

REFERENCES

Gibson, J.P., Freeman, A.E., Boettcher, P.J. (1997) *Livest. Prod. Sci.47:*115-124
Groeneveld, E.(1996) REML VCE User's Guide Version 3.2
Hutchinson, C.A., Newbold, J.E., Potter, S.S., Edgell, M.H.(1974) *Nature* **251**:536-538
Seykora, A.J. and McDaniel, B.T.(1983) *J. Dairy. Sci. 66:* 1486-1493

Estimation of variance of maternal lineage effects at the Langhill dairy herd

T. Roughsedge[1], S. Brotherstone[2] and P. M. Visscher[1]

[1]*Institute of Ecology and Resource Management, University of Edinburgh, Edinburgh EH9 3JG*
[2]*Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh EH9 3JT*

Abstract

Evidence to support the existence of a maternal lineage variance component for production and food intake traits at the Langhill experimental dairy herd was investigated. Maternal pedigree records of the herd were traced back to the points of cytoplasmic origin using herd book records. Cytoplasmic origin was defined as the earliest maternal ancestor of a cow and used to assign cows to maternal lineages. This was either a grade-up cow or an ancestor traced back to 1920. The tracing resulted in the cows being assigned to 56 maternal lineages, ranging in size from one to 72 cows. A total of 1118 records of 517 cows, all with a first lactation record, were used in the analysis. Traits analysed were daily milk, fat and protein yield, fat %, protein %, food dry-matter intake, net energy of milk production, a measure of milk production efficiency, average condition, and calving condition, all averaged over the first 26 weeks of lactation. The analysis was performed using a residual maximum likelihood animal model with and without a random component for maternal lineage. Possible bias, due to the fact that the sires were a select sample from the population, was also examined. No significant effect was found in the analysis of the full data set that could be assigned to maternal lineage. Fat yield was the only trait to show a variance component approaching a 5% significance level with a magnitude of 4% of phenotypic variance. However, when maternal lineages of at least five cows were considered, a significant 4% maternal lineage component of phenotypic variance was found for fat yield. The power of the analysis to detect a variance component of less than 4% was shown to be poor. No evidence was found for a maternal lineage component of food intake traits or condition score. Treating sire as a fixed effect or regressing data on sire EBV made little difference to the maternal lineage component.

Keywords: *cytoplasmic inheritance, dairy cows, heritability, maternal effects, milk yield.*

Introduction

Over recent years it has been hypothesized that a component of the phenotypic variance of economically important traits in dairy cattle is inherited exclusively through the maternal line (e.g. Bell *et al.*, 1985; Schutz *et al.*, 1992). This component of inheritance is of cytoplasmic rather than nuclear origin. It is a long held belief by cattle breeders that some cow families are more important in breeding terms than others, and cows from these families will be used as breeding animals in preference to other cows with similar breeding values. There is also evidence to show that estimates of heritability from daughter dam regressions are higher than those estimates from paternal half-sib analysis (e.g. Seykora and McDaniel, 1983; Visscher and Thompson, 1992). This may suggest that there is a mechanism of inheritance, in addition to nuclear genetic inheritance, that is being transmitted through the female line, which is not being accounted for by current evaluations. A possible explanation for this is the almost exclusively maternal transmission of mitochondrial DNA (mtDNA) in mammalian species (Hutchinson *et al.*, 1974).

Although mtDNA encodes only about 0·05% of all the nuclear plus cytoplasmic genes it is the nature of the potential contribution to phenotypic variation that makes it important. The mitochondria are the energy factories of the cell, and mtDNA codes 13 of

Simm, G., Persaud, P., Neilson, D. R., Parkinson, H. and McGuirk, B. J. 1991. Predicting food intake in dairy heifers from early lactation records. *Animal Production* **52:** 421-434.

Stram, D. O. and Lee, J. W. 1994. Variance-components testing in the longitudinal mixed effects model. *Biometrics* **50:** 1171-1177.

Veerkamp, R. F., Simm, G. and Oldham, J. D. 1994. Effect of interaction between genotype and feeding system on milk production, feed intake, efficiency and body tissue mobilization in dairy cows. *Livestock Production Science* **39:** 229-241.

Visscher, P. M. and Thompson, R. 1992. Comparisons between genetic variances estimated from different types of relatives in dairy cattle. *Animal Production* **55:** 315-320.

Wilmut, I., Schnieke, A. E., McWhir, J., Kind, A. J. and Campbell, K. H. S. 1997. Viable offspring derived from fetal and adult mammalian cells. *Nature* **385:** 810-813.

(*Received* 8 *May* 1998—*Accepted* 5 *August* 1998)

Using pedigree analysis to determine the genetic diversity of the British dairy population over the last thirty years

[1]T. Roughsedge, [2]S. Brotherstone, [1]P.M. Visscher

[1] *Institute of Ecology and Resource Management, University of Edinburgh, Edinburgh EH9 3JG*

[2] *Institute of Cell and Population Biology, University of Edinburgh, Edinburgh EH9 3JT*

Introduction Over the last twenty years the British dairy population has undergone a large importation of Holstein genes, largely from North America and Canada. This study aimed to investigate what effect this importation has had on the genetic diversity of the population and also to look at the change in genetic diversity of the population over the last thirty years.

Materials and methods As computing resources did not facilitate sampling of all cows born in the last thirty years, between 4 and 10 random samples of 2000 Holstein-Friesian cows, born every 5 years from 1967 to 1997 inclusive, were taken from the Holstein Friesian Society of Great Britain and Ireland's database. The base population of this database was born about 1960. The complete ancestry of these cows was traced on the database and these sample pedigrees were used for the various analyses. These ranged from the more traditional determination of inbreeding coefficients and average relationships, to the determination of genetic diversity using techniques outlined by Boichard *et al.* (1997). The diversity measures used were founder equivalent number (FE) and founder ancestor number (FA). FE is defined as the number of equally contributing founders that would be responsible for the current level of genetic diversity in the population and FA considers all ancestors, not just founders, and estimates the number of equally contributing ancestors that would be responsible for the current genetic diversity. The rationale behind FA is that given the heavy use of AI, more recent ancestors with a high usage can cause bottlenecks in genetic diversity. The highest contributing ancestors can also be identified using FA.

Results The results in Table 1 show how the parameters measuring average relationship and inbreeding coefficient were reduced initially by the importation of Holstein genes. However, the actual genetic diversity of the population is in a steady decline that has not been halted by the importation. Indeed it may be suggested that the increased use of fewer sires over this time has reduced the diversity of not only the British population but also the global population of black and white cows.

Table 1 *Parameters of genetic diversity of the British Holstein-Friesian population*

Year	Avg. Inb. Coeff. (%)	Avg. Rel. (%)	Number of Founders	FE	FA
1967	0.03	0.16	2535	1353	702
1972	0.56	0.52	3523	586	276
1977	0.82	0.90	4349	337	175
1982	0.74	1.25	8635	237	169
1987	0.53	1.02	8414	227	163
1992	0.38	1.03	6090	151	144
1997	0.43	1.34	6640	93	93

Table 2 *Most important ancestor contributions*

	1967	1972	1977	1982	1987	1992	1997
First ancestor †	0.023	0.032	0.033	0.026	0.028	0.041	0.051
Second ancestor †	0.017	0.028	0.030	0.025	0.022	0.034	0.046
Third ancestor †	0.007	0.013	0.023	0.024	0.021	0.025	0.030
First 10 ancestors †	0.080	0.140	0.165	0.201	0.185	0.197	0.282
First 50 ancestors †	0.165	0.301	0.359	0.409	0.440	0.427	0.501

† Proportion of genome contributed by.....

Table 2 shows the highest ancestral genomic contributions to the samples taken from the population. It can be seen that fewer animals are contributing more as time passes which is reducing the genetic diversity of the population. For example we see that one bull contributed 5% of the average individual Holstein-Friesian cow genome in 1997.

Conclusion While it is possible to keep the level of inbreeding in the population low by avoiding the mating of closely related individuals, the average relationship of the whole population is steadily increasing as a consequence of the decrease in genetic diversity. The heavy reliance on few selected individuals as sires in the population will inevitably lead to a further decrease in genetic diversity and account of this should be considered in future selection programs.

Acknowledgement Tim Roughsedge thanks MAFF for his funding.

References

Boichard, D., Maignel, L., and Verrier, E. 1997. The value of using probabilities of gene origin to measure genetic variability in a population. *Genetics, Selection, Evolution* **29**:5-23.

Reprinted from

LIVESTOCK PRODUCTION SCIENCE

Livestock Production Science 60 (1999) 359–369

Quantifying genetic contributions to a dairy cattle population using pedigree analysis

T. Roughsedge[a,*], S. Brotherstone[b], P.M. Visscher[a]

[a] *Institute of Ecology and Resource Management, University of Edinburgh, Edinburgh EH9 3JG, UK*
[b] *Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh EH9 3JT, UK*

ELSEVIER

LIVESTOCK PRODUCTION SCIENCE

EUROPEAN ASSOCIATION FOR ANIMAL PRODUCTION

The object of the Association is to promote sound development of livestock production by the international exchange and synthesis of research results. The journal *Livestock Production Science* intends to support these aims by publication of original research papers and comprehensive reviews in the field of livestock production and by providing a medium for the exchange of current views. It is hoped as a result to strengthen the contacts between the experts and organizations of different countries.

Methods and objectives of reported animal experimentation, aiming at the advancement of knowledge and of production efficiency, should meet ethical standards of animal treatment, and a justification of adverse conditions is required.

Publication information: *Livestock Production Science* (ISSN 0301-6226). For 1999 volumes 57–61 are scheduled for publication. Subscription prices are available upon request from the Publisher. Subscriptions are accepted on a prepaid basis only and are entered on a calendar year basis. Issues are sent by surface mail except to the following countries where Air delivery via SAL mail is ensured: Argentina, Australia, Brazil, Canada, Hong Kong, India, Israel, Japan, Malaysia, Mexico, New Zealand, Pakistan, PR China, Singapore, South Africa, South Korea, Taiwan, Thailand, USA. For all other countries airmail rates are available upon request. Claims for missing issues should be made within six months of our publication (mailing) date.
Orders, claims, and product enquiries: please contact the Customer Support Department at the Regional Sales Office nearest you:
New York: Elsevier Science, P.O. Box 945, New York, NY 10159-0945, USA. Tel. (+1) 212-633-3730, [Toll free number for North American customers: 1-888-4ES-INFO (437-4636)], Fax (+1) 212-633-3680, E-mail usinfo-f@elsevier.com
Amsterdam: Elsevier Science, P.O. Box 211, 1000 AE Amsterdam, The Netherlands. Tel. (+31) 20-485-3757, Fax (+31) 20-485-3432, E-mail nlinfo-f@elsevier.nl
Tokyo: Elsevier Science K.K., 9-15 Higashi-Azabu 1-chome, Minato-ku, Tokyo 106-0044, Japan. Tel. (+81) 3-5561-5033, Fax (+81) 3-5561-5047, E-mail info@elsevier.co.jp
Singapore: Elsevier Science, No. 1 Temasek Avenue, #17-01 Millenia Tower, Singapore 039192. Tel. (+65) 434-3727, Fax (+65) 337-2230, E-mail asiainfo@elsevier.com.sg
Rio de Janeiro: Elsevier Science, Rua Sete de Setembro 111/16 Andar, 20050-002 Centro, Rio de Janerio – RJ, Brazil; phone: (+55) (21) 509 5340; fax: (+55) (21) 507 1991; e-mail; elsevier@campus.com.br [Note (Latin America): for orders, claims and help desk information, please contact the Regional Sales Office in New York as listed above]

Back volumes: Please contact the Publisher.

ELSEVIER

Livestock Production Science 60 (1999) 359–369

LIVESTOCK PRODUCTION SCIENCE

Quantifying genetic contributions to a dairy cattle population using pedigree analysis

T. Roughsedge[a,*], S. Brotherstone[b], P.M. Visscher[a]

[a]*Institute of Ecology and Resource Management, University of Edinburgh, Edinburgh EH9 3JG, UK*
[b]*Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh EH9 3JT, UK*

Abstract

A variety of techniques were employed in the analysis of the change in genetic diversity of the UK Holstein–Friesian population over the last 30 years using the Holstein Friesian Society database (which has a base year of 1960). The parameters estimated were average inbreeding coefficient, average degree of relationship between cows, and two measures of genetic diversity, founder equivalent and founder ancestor number. The cow population was seen to change in founder origin from 96% British Friesian in 1967 to 24% British Friesian and 76% North American Holstein in 1997. The change in origin was seen to affect the rate of increase in inbreeding and to a lesser extent relationship, however the measures of genetic diversity were largely unaffected by the Holstein importation. In 1997 average relationship between cows had reached 1.34%, average inbreeding coefficient was 0.4% and the founder equivalent and founder ancestor number had converged at 93. The average inbreeding coefficient was seen to fall from 0.74% in 1982 to 0.38% in 1992 and to remain fairly constant up to 1997. The maternal structure of the cow population born in 1997 was also analysed. It was found that 93% of the cows were in maternal families of only one to four cows and only 0.5% of cows were in maternal families with more than 100 members, where a maternal family is a group of cows related only by maternal lineage.

Keywords: Holstein–Friesian; Pedigree; Inbreeding; Relationship; Founder

1. Introduction

Over the last 20 years there has been a rapid increase in the proportion of North American Holstein genes in the British black and white dairy cattle population. More recently genetics from European countries such as France and the Netherlands have been imported. However, these countries themselves have sourced most of their genetics from North America, so that we do not distinguish between North American and European genes in this study. This is not the first time that an importation has occurred in the UK dairy population. Robertson and Asker (1951) provided a brief history of the formation of the British Friesian herd tracing the effects that a variety of different importations had had. They document importations that occurred from Holland in the last century and the early part of this century and also a Canadian importation in 1946. The recent importation has certainly been the most significant

*Corresponding author.

0301-6226/99/$ – see front matter
PII: S0301-6226(99)00106-2

due to the advances that have occurred in reproductive technology allowing a more rapid and widespread grading up of animals. Within this study the British Friesian population can be defined as those individuals registered with a British Friesian breed code in the database and foreign animals can be defined as other breed codes, those individuals being mostly of North American Holstein origin. The breed substitution has had an effect on the degree of relationship and level of inbreeding in the British dairy population, and the extent of this was investigated. Inbreeding trend has for many years been the preferred tool for the quantification of the rate of genetic drift but, as highlighted by Boichard et al. (1997), in order for the measure to be of relevance the population should conform to the rules specified by Wright (1931). It should be closed, unselected, panmictic and of finite size. It is quite obvious that the British dairy population does not satisfy these requirements. Boichard et al. (1997) proposed the use of measures of genetic variability more commonly used in conservation genetics. Lacy (1989) developed a 'founder equivalent' approach which combined the information of the founder animals contributing to the population under study and estimated the number of equally contributing ancestors that would provide the same level of genetic diversity. Caballero (personal communication) indicated that this parameter is directly related to group coancestry (Cockerham, 1967), the average pairwise coancestry of a group of individuals in a pedigree including reciprocals and self-coancestries. The parameter can also be related to genetic diversity defined as the expected frequency of heterozygotes by descent. This approach was taken further by Boichard et al. (1997) in the development of 'founder ancestor equivalent' which located the highest contributing ancestors to the population under study and used these animals to estimate the equivalent number of equally contributing ancestors that would provide the same level of genetic diversity. Bowman et al. (1978) studied the degree of relationship and average level of inbreeding in the British Friesian population. This current study is intended to update the information on the British dairy population and to investigate other properties of the dairy population by estimating parameters of genetic diversity.

2. Material and methods

2.1. Data

Data were extracted from the database of the Holstein Friesian Society of Great Britain and Ireland which has complete pedigree records, including grade-up animals, back to 1960. In this study grade-up animals are the progeny of non-pedigree cows that have been mated to pedigree bulls. This database also records the pedigree of all imported foreign cows and, if available, three generations of ancestors. In total the database holds about 5 million animals. Random samples, each of 2000 cows, were taken from cows born in specific years and the complete ancestry of these animals present in the database was traced. This pedigree file was then used for the various analyses. Ten samples were taken for 1997 births, which was essentially the year of interest. A further four samples per year were then taken going back in 5-year intervals including 1992, 1987, 1982, 1977, 1972, 1967. Five years was taken as being approximately the equivalent of a generation. The use of sampling rather than analysing the whole population was necessitated by a restriction on computing resources.

2.2. Founder equivalent

Founder animals were taken as being ancestors having unknown parents. If only one parent was unknown the animal was treated as a half founder and its contribution to the population under consideration was halved. The formula for the calculation of the founder equivalent (FE) parameter is:

$$\mathrm{FE} = 1 \Big/ \sum_{k=1}^{f} q_k^2$$

where q is the proportional contribution of a founder k (total number of founders is f), to the generation for which FE is being calculated. If all founders had contributed equally to the population then the FE would equal the actual number of founders. This provides a measure of the change in genetic diversity of the population over time.

2.3. Founder ancestor number

The FE parameter does not properly account for bottlenecks in the pedigree caused by the heavy use of AI in cattle populations. If a recent sire made a high contribution to the current population through AI then it would be more important, in respect to the loss of genetic diversity, than tracing all of the pathways that founder ancestors in the pedigree made through this individual. In order to account for this Boichard et al. (1997) proposed a second measure, founder ancestor number (FA), which treats all the ancestors in the same way regardless of whether they are founders. The highest contributing ancestor, termed a 'pseudo-founder' was identified, its contribution stored, and then it was removed from the pedigree so that no further contribution could be made through its pathway in the pedigree. The procedure was iterated, finding the highest marginal contribution at each cycle. At each round of iteration an upper and lower limit was imposed on FA. The upper limit (f_u) was determined by setting the remaining contribution to the population genome to be equally distributed across all remaining founders and calculating FA. The lower limit (f_l) was determined by considering that all remaining founders contribute the proportion of the genome equivalent to the last found pseudo founder. The upper and lower limits converge over iterations and a criterion for stopping can be imposed on this process using either a level of convergence or a given number of pseudo founders. In this study a stopping criteria of $f_u - f_l < 5$ was used.

2.4. Founder origin

The contribution of the founder animals to the average individual genome of the population was determined in order to show the influence that the importation of the North American genome has had over the last 30 years. This involved finding the earliest known ancestors of the cows in the sample and calculating their contribution to the current gene pool based on each parent contributing 50% of their offspring's genome. The information generated by the FE procedure was used for this analysis.

2.5. Inbreeding coefficient and average relationship

Inbreeding coefficient was determined using the methodology of Meuwissen and Luo (1992). This was based on the decomposition of the additive genetic relationship matrix **A**, as described by Henderson (1976): $\mathbf{A}=\mathbf{LDL}'$, where **L** is a lower triangular matrix containing the fraction of genes that animals derive from their ancestors, and **D** is a diagonal matrix containing the within family additive genetic variances of animals. **L** is calculated row by row in the algorithm. The results of the samples for a given year were pooled. Average coefficient of inbreeding was determined for the whole population and also for those individuals with a non-zero coefficient of inbreeding. The average relationship between all members of a given sample, born in the year of the sample, was determined using the recursive algorithm of Miglior et al. (1992).

2.6. Maternal family distribution

In addition to the nuclear genome the current distribution of the mitochondrial genome in the Holstein Friesian cow population was analysed using all cows born in 1997 as the study population. The earliest maternal ancestor for each of the Holstein Friesian cows born in 1997 was traced using the same database as the other analyses and this was taken as being the point of cytoplasmic origin. Mitochondrial DNA (mtDNA) is inherited almost exclusively maternally in mammalian species (Hutchinson et al., 1974) and as such all cows in a maternal family will have the same mtDNA as the cow that provides the cytoplasmic origin of that maternal family.

3. Results

3.1. Samples

For all parameters estimated the samples provided consistent results within the year sampled. For example in 1997 the mean and standard deviation of the sample results was 0.43 ± 0.057 for average

inbreeding coefficient, 1.34±0.09 for average degree of relationship and 93±4.26 for FE. The results presented are a mean of the sample results for each year. Table 1 gives details of the sample pedigrees and shows that the 'depth' of pedigree information changed in the early 1990's indicative of the change to North American Holstein genome origin, which has less ancestral information stored on the database.

3.2. Founder origin

The change in the origin of the average genome over time (Table 2) clearly demonstrates the influence of the North American Holstein breed on the British Friesian population. The change in origin was most clearly seen over the last 10 years when male Holstein origin rose to the same level of influence as British Friesian females had had 25 years ago. Holstein female origin had risen to 20% in 1997 but the rate of increase was seen to fall to only 2% (0.4%/year) between 1992 and 1997. In all four origins, except for the Holstein female, the rate of change did not appear to have stabilised by 1997. It is important to note that the completeness of the pedigree information has an influence on the male:female contribution ratio. For example consider a simple scenario of a cow with pedigree information available on the sire and dam and the paternal grandsire and granddam only. If base animals are taken as being animals with unknown parents then 75% (50% dam, 25% paternal granddam) of the genome is traced to female origin and 25% (25% paternal grandsire) to male origin.

3.3. Average inbreeding coefficient

The base population of the pedigree samples used in this analysis was 1960. The samples from 1967 were, therefore, not very deep in pedigree structure (Table 1) and, as can be seen in Fig. 1 and Table 3, calculated inbreeding had risen little above zero at this time. Bowman et al. (1978) calculated average inbreeding coefficient of cows in the British Friesian population between 1955 and 1972, tracing animals

Table 1
Average information about pedigrees for 10 samples of 2000 cows

Year of birth	Number of founders in sample pedigree	Number of ancestors in sample pedigree	Average number of ancestors per individual
1967	2535	2785	2
1972	3523	7492	53
1977	4349	11 842	254
1982	8635	33 210	872
1987	8414	31 361	1321
1992	6090	22 284	796
1997	6640[a]	26 434[b]	646

[a] SD between 1997 samples 134.
[b] SD between 1997 samples 454.

Table 2
Origin of nuclear genome of the British Holstein–Friesian population

Year of birth	British female (%)	Male (%)	Foreign female (%)	Male (%)
1967	72	24	2	2
1972	55	37	4	4
1977	53	34	6	7
1982	48	31	9	12
1987	36	28	14	22
1992	22	20	18	40
1997	13	11	20	56

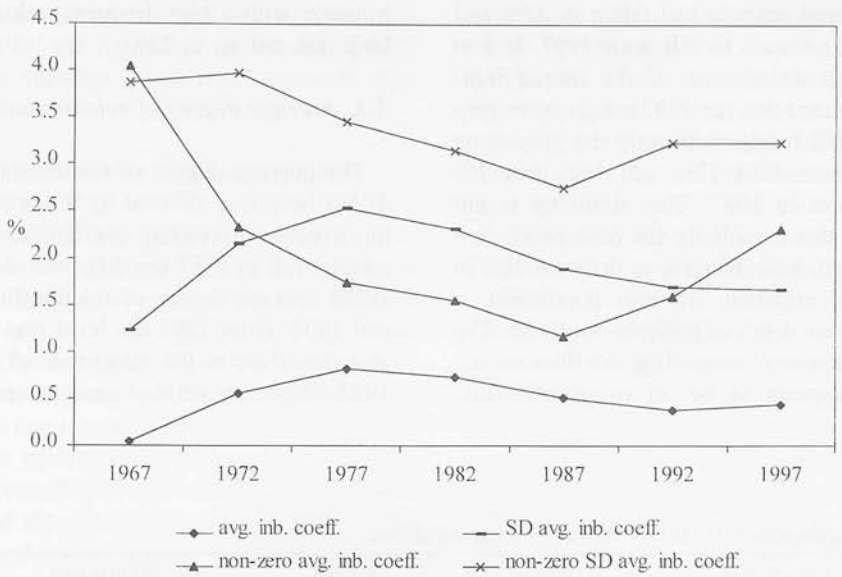

Fig. 1. Average and non-zero average inbreeding coefficient of cows of the UK dairy population born between 1967–1997.

back to the 1914 herd book. They used the sampling techniques described by Wright and McPhee (1925), that involve using three generations of complete pedigree information and then a sampling technique to establish common ancestors prior to these generations. In 1960, the base year in this current study, they estimated an average coefficient of inbreeding of 2.28%. The rate of increase in average coefficient of inbreeding in the study of Bowman et al. (1978) between 1965 and 1972 was very similar to that seen in the current study in which the level of inbreeding in the population followed a fairly steady increase between 1967 and 1982, over which time the origins of the population were in a steady state (Table 2). However, between 1982 and 1987 a steady fall was seen in the average inbreeding coefficient and this fall continued until 1992. From 1992 to 1997 a slight rise of 0.05% was seen, though when the years between 1992 and 1997 were analysed (results not shown) the samples were seen to differ little from 0.4%. More interesting were the results of the non-zero average inbreeding coefficients and the distribution of the inbred individuals. The year 1967 can be discounted as being too close to the base population to have had a significant level of inbreeding with only 1% of the population inbred, but by 1972 24% of the population was inbred and this rose to nearly 50% in 1977 and 1982. By 1987 the

Table 3
Average inbreeding coefficient (%) of the British Holstein–Friesian population[a]

Year of birth	Avg. inb. coeff.	SD avg. inb. coeff.	Non-zero avg. inb. coeff.	SD non-zero avg. inb. coeff.	Proportion inbred (%)
1967	0.05	1.22	4.05	3.87	1
1972	0.55	2.16	2.33	3.97	24
1977	0.82	2.54	1.75	3.46	47
1982	0.74	2.31	1.55	3.15	48
1987	0.52	1.90	1.17	2.76	42
1992	0.38	1.70	1.60	3.24	23
1997	0.43	1.67	2.32	3.24	19

[a] Base population 1960.

proportion of inbred animals had fallen to 42% and this proportion continued to fall until 1997. It was evident when the distribution of the inbred individuals was examined that the 1997 results were very similar to the 1972 results with only the proportion of individuals more than 12% and less than 1% inbred being lower in 1997. This similarity is not surprising given that essentially the base population of the 1997 animals had changed, as demonstrated in Table 2, to the imported Holstein population, a population with less depth of pedigree available. The distribution of non-zero inbreeding coefficients across all years appears to be an exponential distribution with a high frequency close to zero and a long thin tail up to 25%.

3.4. Average degree of relationship

The average degree of relationship (Fig. 2, Table 4) has been less affected by the breed grade-up than the average inbreeding coefficient. Though the parameter fell in 1987 the drop was soon halted with a stable average degree of relationship between 1987 and 1992. After 1992 the level was seen to increase at a rate close to the magnitude of the rate prior to 1982. However, what is more interesting is that the

Table 4
Average pair-wise relationship of the British Holstein–Friesian population

Year of birth	Avg. rel. (%) (%)	SD avg. rel.	Non-zero avg. rel. (%)	SD non-zero avg. rel.	Non-zero proportion (%)
1967	0.16	3.94	1.83	1.17	9
1972	0.52	2.62	1.08	1.90	48
1977	0.90	2.46	1.14	2.22	79
1982	1.25	2.64	1.46	2.50	85
1987	1.02	2.47	1.28	2.26	80
1992	1.03	2.94	1.52	2.50	60
1997	1.34	2.87	1.80	2.59	93

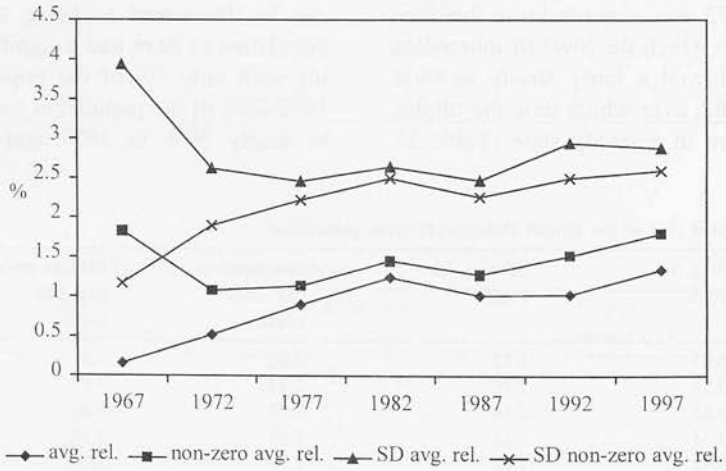

Fig. 2. Average and non-zero average degree of relationship between cows in the UK Holstein–Friesian population born between 1967 and 1997.

proportion of inbred individuals fell in 1987 and 1992 but by 1997 the figure had risen, with 93% of the population showing a non-zero degree of relationship.

3.5. Founder and ancestor equivalent number

The FE and FA parameter estimates for the samples (Table 5) illustrate the fall in genetic diversity over the last 30 years. The values shown for FA were actually the upper limits given the convergence criterion that $f_u - f_l >= 5$, and therefore potentially over-estimated by five. The trend in FE and FA shows that the fall in genetic diversity of the population was greatest in the years close to the base population. From 1977 to 1987 the FE parameter fell very little and the FA parameter, which estimated variability based on the highest contributing ancestors remained constant from 1977 until 1992. In 1997 both parameters had fallen to the same value of about 90. The convergence of these parameters in 1997 is indicative of the change in the founder population to the imported Holstein population with few selected sires being used to grade the population up. These few, but highly influential, founder individuals were also identified as being the highest contributing ancestors and hence the two parameters were seen to converge. To further illustrate the effect of few highly selected sires being extensively used, Table 6 shows the highest ancestor contributions to the years investigated. The highest contributing ancestor in 1997, To-Mar Blackstar, was responsible for 5% of the average nuclear genome in the population, more than double the level seen in 1967. Hanoverhill Starbuck made the second highest contribution, of 4.6% of the nuclear genome, in 1997, three times the contribution of the second highest contributing ancestor in 1967. The top 50 ancestors in 1997 were responsible for 50% of the nuclear genome, more than three times the 1967 level.

Table 5
Parameters of genetic diversity of the British Holstein–Friesian population[a,b]

Year of birth	Number of founders	FE	FA	FE as proportion of founders (%)	FA as proportion of founders (%)
1967	2535	1353	702	53	28
1972	3523	586	276	17	8
1977	4349	337	175	8	4
1982	8635	237	169	3	2
1987	8414	227	163	3	2
1992	6090	151	144	2	2
1997	6640	93	93	1	1

[a] FA is the founder ancestor equivalent number.
[b] FE is the founder equivalent number.

Table 6
Most important ancestor contributions

	1967	1972	1977	1982	1987	1992	1997
First ancestor[a]	0.023	0.032	0.033	0.026	0.028	0.041	0.051
Second ancestor[a]	0.017	0.028	0.030	0.025	0.022	0.034	0.046
Third ancestor[a]	0.007	0.013	0.023	0.024	0.021	0.025	0.030
First 10 ancestors[a]	0.080	0.140	0.165	0.201	0.185	0.197	0.282
First 50 ancestors[a]	0.165	0.301	0.359	0.409	0.440	0.427	0.501

[a] Proportion of alleles contributed by . . . . .

Table 7
Distribution of Holstein–Friesian cows born in 1997, across maternal families

Family size (no. of cows)	No. of families	No. of cows	Proportion of cows (%)	Average no. of generations to origin
1	69 575	69 575	32.1	3.3
2–5	36 009	96 212	44.4	4.7
6–10	3743	27 277	12.6	6.0
11–20	1017	13 986	6.5	6.4
21–30	150	3721	1.6	6.8
31–40	56	1932	0.9	6.8
41–100	49	3046	1.4	7.4
101–150	5	574	0.3	7.7
151–220	2	383	0.2	9.0
Total	110 607	216 706	100	7.6

3.6. Maternal family distribution

The distribution of cows across maternal families can be seen in Table 7. Most of the cows are seen in families of less than five cows. The average family size was two with a standard deviation of 2.75 and nearly 70% of the families had only one cow. These single cow families represent 32% of the registered cows born in 1997 and traced on average three generations to their cytoplasmic origin, with about 70% tracing to grade-up cows. When the range of family size was increased to families of one to four cows 93% of the families were accounted for containing 71% of the cows. Only seven families were greater than 100 cows, with one having 217 members.

4. Discussion

In this study it has been shown that by treating the Holstein population as unrelated to the Friesian population, a decrease in inbreeding level was seen over the last 10 years. The same trend was also evident in the Dutch black and white dairy population between 1980 and 1985 (Te Braake et. al., 1994). Te Braake et. al. used a similar methodology in the calculation of average inbreeding coefficient and the population they were studying had undergone a similar introgression of North American Holstein genomes. Given the proportion of the British black and white cattle genome that is Holstein in origin at the present time, the relationship between the two breeds will soon be of little relevance. If this logic is followed it is not unreasonable to hypothesise that the rate of increase in the average level of inbreeding will soon return to the rate seen prior to the Holstein importation or indeed to a higher rate. In comparison, work done estimating inbreeding coefficients in the United States (VanRaden, 1992; Wiggans et al., 1995) and Canada (Miglior and Burnside, 1995) has consistently shown higher levels of inbreeding than has been seen in the British Holstein–Friesian population. VanRaden (1992) and Wiggans et al. (1995) using the same population of US Holsteins with a 1960 base population calculated the average inbreeding coefficients as 0.4, 1, 2 and 2.6%, respectively, for 1970, 1980, 1987 and 1990. Miglior and Burnside (1995) using a base population of 1950 estimated an average inbreeding coefficient of 1.7% between 1986 and 1990. They showed a plateau from the late 1970's until the late 1980's due to the heavy use of US Holstein sires in the Canadian population.

The average degree of relationship parameter would appear to have been less sensitive to the introduction of Holstein genes than the level of inbreeding. Indeed given the steady increase in average relationship and the introgression of North American Holsteins across most of Europe, it is

inevitable that the average inbreeding coefficient will increase in the future. In the study of Bowman et al. (1978) the degree of relationship estimated for 1960 was 1.45% but was seen to have risen to 3.19% by 1972, with the rate of increase between 1965 and 1972 almost identical to that estimated by this current study.

As has previously been discussed the recent change in breed origin resulted in both the rate of increase in average inbreeding coefficient and average degree of relationship to be lowered considerably and indeed a decrease in the inbreeding parameter was seen during the late 1980s and early 1990s. These parameters are not therefore able to provide a good estimate of the genetic diversity of the British black and white dairy population over recent years. FE and FA were used to illustrated the change in genetic diversity over time, whilst the population was undergoing a major grade-up, which was not possible using the average inbreeding coefficient. These parameters provided a useful means of describing the change in genetic diversity but as indicated by Boichard et al. (1997) they cannot be used as predictors of future variability. In previous studies using FE and FA (Boichard et al., 1997; Sölkner et al., 1998) whole populations have been used. In this study samples from the population were used and whilst the parameters provided a description of the change over time they did not necessarily accurately estimate the magnitude of the population parameters. However, given the nature of these parameters, i.e. that they are heavily dependent on the proportional contribution of the top few, highest contributing founders/ancestors, it can be argued that the estimates provide a close approximation. The identity and magnitude of the contribution, of the highest contributing ancestors, was found to be consistent across the samples for given years, and this further validated the accuracy of the sample result. Young and Seykora (1996) estimated the ancestors with the highest contribution to the 1990 US Holstein females. Two of the three bulls with the highest contribution to the US female genome in 1990 were the same two bulls that made the highest contribution to the British Holstein–Friesian female genome in 1992. Round Oak Rag Apple Elevation contributed 12.2 and 3.4% and SWD Valiant contributed 9.6 and 4.1% to the US and UK female populations, respectively.

The use of long-term founder contributions is similar to the approach taken by Woolliams and Mantysaari (1995) in the Finnish Ayrshire population, however they related this contribution to the rate of inbreeding using the methodology of Wray and Thompson (1990).

The use of founder equivalent as a measure of genetic diversity can be criticised due to the nature of long-term founder contributions. Over a number of generations the proportionate contributions of founder individuals will stabilise to the same across all contemporary individuals (Wray et al., 1994) and hence the founder equivalent parameter will become fixed. Caballero (personal communication) states that minimising coancestries is a more effective way of maintaining genetic variability than trying to equalise founder contributions. This minimises the variances of contributions from ancestors to descendants in all previous generations to the current one, i.e. maximising the effective population size.

Rate of increase in average inbreeding coefficient has been used frequently as a measure of genetic diversity (e.g. Goddard, 1992; Wang, 1997). However, it is a measure that is sensitive to changes in population structure and crossbreeding. One method to avoid this problem was proposed by VanRaden (1992). He suggested that given that Holstein and Friesian populations will have a common base at some point in the past, then it is possible to assign an average relationship between Friesian and Holstein individuals based on several assumptions utilising earliest known ancestor information. The genetic diversity parameter FA estimated in this study is less sensitive to a change in origin and reflects the level of contribution being made by few very influential individuals to the whole population. Such information could also be used to identify families that have high influence and provide extra information to be used in future selection strategies. These measures of founder representation are used by conservation biologists in order to retain genetic variation of captivity bred 'wild' populations (Lacy, 1989). As FE and FA decrease so the level of inbreeding will increase with the general increase in similarity of the genome. The present study estimated the genetic

diversity parameters to be of very similar magnitude to those estimated by Boichard et al. (1997) and Sölkner et al. (1998) for dairy breeds. However, the FA and FE parameters converged in 1997 in the British Holstein–Friesian population. This was attributable to the fact that the imported Holstein sires formed the new founder base population and, at the same time, these few, heavily used sires made a large contribution to the average genome of the population. This again illustrates the robustness of the FA parameter, which does not rely on distant relationships to the same level that the other parameters do. The measure of FE in small conservation populations is used as a standard by which genetic diversity can be maintained. Here the populations are not being actively selected and the objective is to attain an equal representation of all founders across generations. In livestock improvement the objective is very different but if we are looking towards long term effective improvement then preservation of the variability of the genetic base is important. Goddard (1992), considering the global black and white population and using N_e as the measure of genetic diversity calculated the optimal effective size using discounting to estimate future gains. Such a calculation is dependent on several assumptions regarding objectives and differences between countries. Goddard (1992) hypothesised that differences between countries in both environment and objectives will lead to the development of isolated strains, which originate from the same founding animals, i.e. North American Holsteins. What is not certain at present is the degree of exchange of genes that will continue to occur between these 'isolated' populations.

When the mitochondrial genome was considered, using 1960 as the population of cytoplasmic origin, a very different level of diversity was seen. Due to the nature of the transmission of the mitochondrial genome, i.e. it is almost exclusively maternally inherited; the heavy use of foreign sires has had no impact on its diversity. This study provided an under estimation of maternal family size for two reasons. Firstly, pedigree information prior to 1960 was not used and secondly many of the single cow families were a first generation grade-up cow with no information on the non-pedigree dam. It would appear that the level of diversity remaining between mitochondrial genomes is still high or at least has not been reduced greatly by selection over the last 30 years.

Acknowledgements

TR acknowledges MAFF for his funding and studentship support. We thank the Holstein–Friesian Society for the use of their database. Bill Hill and John Woolliams are thanked for constructive discussions. We would also like to thank the reviewer for very helpful comments.

References

Boichard, D., Maignel, L., Verrier, E., 1997. The value of using probabilities of gene origin to measure genetic variability in a population. Genet. Select. Evolut. 29, 5–23.

Bowman, J.C., Butler, E.A., Tuncel, E., 1978. Coefficients of inbreeding and degree of relationship for the British Friesian herd. Anim. Prod. 27, 269–276.

Cockerham, C.C., 1967. Group inbreeding and coancestry. Genetics 56, 89–104.

Goddard, M.E., 1992. Optimal effective population size for the global population of black and white dairy cattle. J. Dairy Sci. 75, 2902–2911.

Henderson, C.R., 1976. A simple method for computing the inverse of a numerator relationship matrix used in the prediction of breeding values. Biometrics 32, 69–83.

Hutchinson, C.A., Newbold, J.E., Potter, S.S., Edgell, M.H., 1974. Maternal inheritance of mammalian mitochondrial DNA. Nature 251, 536–538.

Lacy, R.C., 1989. Analysis of founder representation in pedigrees: founder equivalents and founder genome equivalents. Zoo Biol. 8, 111–123.

Meuwissen, T.H.E., Luo, Z., 1992. Computing inbreeding coefficients in large populations. Genet. Select. Evolut. 24, 305–312.

Miglior, F., Szkotnicki, B., Burnside, E.B., 1992. Analysis of levels of inbreeding and inbreeding depression in Jersey cattle. J. Dairy Sci. 75, 1112–1118.

Miglior, F., Burnside, E.B., 1995. Inbreeding of Canadian Holstein cattle. J. Dairy Sci. 78, 1163–1167.

Robertson, A., Asker, A.A., 1951. The genetic history and breed structure of British Friesian cattle. Empire J. Exp. Agric. 19, 113–130.

Sölkner, J., Filipcic, L., Hampshire, N., 1998. Genetic variability of populations and similarity of subpopulations in Austrian cattle breeds determined by analysis of pedigrees. Anim. Sci. 67, 249–256.

Te Braake, M.F.H., Groen, A.F., Van Der Lugt, A.W., 1994. Trends in inbreeding in Dutch Black and White dairy cattle. J. Anim. Breed. Genet. 111, 356–366.

VanRaden, P.M., 1992. Accounting for inbreeding and crossbreeding in genetic evaluation of large populations. J. Dairy Sci. 75, 3136–3144.

Wang, J., 1997. More efficient breeding systems for controlling inbreeding and effective size in animal populations. Heredity 79, 591–599.

Wiggans, G.R., Van Raden, P.M., Zuurbier, J., 1995. Calculation and use of inbreeding coefficients for genetic evaluation of United States Dairy Cattle. J. Dairy Sci. 78, 1584–1590.

Woolliams, J.A., Mantysaari, E.A., 1995. Genetic contributions of Finnish Ayrshire bulls over four generations. Anim. Sci. 61, 177–187.

Wray, N.R., Thompson, R., 1990. Prediction of rates of inbreeding in selected populations. Genetical Res. 55, 41–54.

Wray, N.R., Woolliams, J.A., Thompson, R., 1994. Prediction of rates of inbreeding in populations undergoing index selection. Theor. Appl. Genet. 87, 878–892.

Wright, S., 1931. Evolution in Mendelian populations. Genetics 16, 97–159.

Wright, S., McPhee, H.C., 1925. An approximate method of calculating coefficients of inbreeding and relationship from livestock pedigrees. J. Agric. Res. 31, 377–383.

Young, C.W., Seykora, A.J., 1996. Estimates of inbreeding and relationship among registered Holstein females in the United States. J. Dairy Sci. 79, 502–505.

LIVESTOCK PRODUCTION SCIENCE

Note for Authors

Types of paper published in the journal
– papers reporting results of original research – review articles – short communications – (guest) editorials – book reviews – news and announcements.

Authors in Japan please note: Upon request, Elsevier Science Japan will provide authors with a list of people who can check and improve the English of their paper (*before submission*). Please contact our Tokyo office: Elsevier Science K.K., 1-9-15 Higashi-Azabu, Minato-ku, Tokyo 106-0044, Japan; Tel.: (+81)-3 5561 5032; Fax: (+81)-3 5561 5045.

Submission of articles: Contributions to the journal (original ms plus 3 copies) should be sent to the Editorial Secretariat, *Livestock Production Science*, P.O. Box 181, 1000 AD Amsterdam, The Netherlands.

All contributions will be carefully refereed for international relevance and quality. Submission of an article is understood to imply that the article is original and unpublished and is not being considered for publication elsewhere.

Electronic manuscripts: Electronic manuscripts have the advantage that there is no need for the rekeying of text, thereby avoiding the possibility of introducing errors and resulting in reliable and fast delivery of proofs. For the initial submission of manuscripts for consideration, hardcopies are sufficient. For the processing of **accepted papers**, electronic versions are preferred. After **final acceptance**, your disk plus two, final and exactly matching printed versions should be submitted together. Double density (DD) or high density (HD) diskettes (3.50 or 5.25 inch) are acceptable. It is important that the file saved is in the native format of the wordprocessor program used. Label the disk with the name of the computer and wordprocessing package used, your name, and the name of the file on the disk. Further information may be obtained from the Publisher.

Enquiries concerning manuscripts and proofs: Questions arising after acceptance of the manuscript, especially those relating to proofs, should be directed to, Elsevier Science Ireland Ltd., Elsevier House, Brookvale Plaza, East Park, Shannon, Co. Clare, Ireland. Tel.: (+353) 61 709640, Fax: (+353) 61 709107.

Proofs: One set of proofs will be sent to the author to be checked for printer's errors. In the case of two or more authors, please indicate to whom the proofs should be sent. Fifty offprints of each article published will be supplied free of charge. Additional offprints can be ordered on an offprint order form which is sent upon receipt of the manuscript by the Publisher.

For a full and complete Guide for Authors please refer to: *Livestock Production Science*, 57 (1998) 91–94.
The instructions can also be found on the World Wide Web: access under
http://www.elsevier.nl/locate/livprodsci or http://www.elsevier.com/locate/livprodsci

***Livestock Production Science* has no page charges**

0301-6226/99/$20.00

♾ The paper used in this publication meets the requirements of ANSI/NISO Z39.48-1992 (Permanence of Paper)

PRINTED IN THE NETHERLANDS

US mailing notice. Livestock Production Science (0301-6226) is published monthly by Elsevier Science (Molenwerf 1, Postbus 211, 1000 AE Amsterdam). Annual subscription price in the USA is US$ 1089.00 (valid in North, Central and South America), including air speed delivery. Periodicals postage rate is paid at Jamaica, NY 11431.
USA POSTMASTERS: Send address changes to *Livestock Production Science*, Publications Expediting, Inc., 200 Meacham Avenue, Elmont, NY 11003. Airfreight and mailing in the USA by Publication Expediting.